THE UNIVERSITY *of* EDINBURGH

Database mining studies on protein-peptide and protein-protein interactions

A Thesis

Submitted for the Degree of

Doctor of Philosophy

by

Calum Stevenson, B.Sc. (Hons)

Structural Biochemistry Group

Institute of Structural and Molecular Biology

University of Edinburgh

2012

Summary of Contents

Title i

Summary of Contents ii

Abstract iii

Declaration v

Acknowledgements vi

List of Abbreviations vii

Table of Contents ix

List of Figures xiv

List of Tables xviii

Chapter 1. Utilising virtual screening to aid the discovery of small molecule ligands to disrupt protein-protein and protein-peptide interactions 1

Chapter 2. Computational and biochemical protocols 32

Chapter 3. Building a model of the GPCR Gonadotrpin Releasing Hormone Receptor (GnRH-R). 58

Chapter 4. Theoretical and experimental studies on GnRH and agonist and antagonist peptide structures 91

Chapter 5. Identification of small molecule inhibitors to GnRH-R using similarity searching 130

Chapter 6. Identifying small molecule inhibitors using a LIDAEUS database mining approach 157

Chapter 7. Biochemical analysis and structural characterisation of Post Synaptic Density protein 95 (PSD95) 177

Chapter 8. Biochemical and structural characterisation of B Lymphocyte Stimulating Factor (BLyS) 221

Abstract

A major area of interest is the identification of proteins that play a role in hormone dependent cancers and in collaboration with the MRC Centre for Reproductive Health we studied the gonadotropin releasing hormone receptor (GnRH-R). Other targets described in the thesis are the SH3 domain of PSD-95 and the protein BLyS.

In order to identify potential inhibitory small molecules we have used a variety of computational data base mining approaches as well as using and developing experimental binding assays. It has become increasingly challenging to evaluate the most representative drug like small molecule compounds when using traditional high throughput screening methods. This thesis assesses the use of *in silico* tools to probe key protein-protein and protein-peptide interactions. These tools provide a means to identify enriched compound datasets which can be purchased and tested *in vitro* in a time and cost efficient way.

The transmembrane protein GnRH-R provides an interesting opportunity to identify small molecules that could inhibit the binding of its peptide ligand GnRH. This is a challenging project as there are few examples in the literature of drug-like molecules that bind to such protein-peptide interfaces. The first step involved receptor modelling using solved crystal structures of homologous proteins. The model was then validated by developing structure activity relationships for established high affinity ligands. We also performed crystallographic and biophysical studies on the native GnRH decapeptide.

Two other protein-protein systems were also examined using the same virtual screening and experimental ligand binding methodology. SH3 domains play an

important role in cell signalling and we used the PSD-95 protein as our target for study as a crystal structure has been published. As well as identifying potential ligands we characterised structural properties of PSD-95 fusion proteins and also developed the basis for compound assay. The third system studied was B Lymphocyte Stimulator (BLyS) which is a target for treatment of a number of autoimmune diseases. This presented an interesting target for study as the protein binds to multiple receptors depending on its multimeric state. BLyS protein was characterised using electron microscopy and other biophysical techniques.

Declaration

The work presented in this thesis is the original work of the author. This thesis has been composed by the author and not been submitted in whole or part for any other degree.

Calum Stevenson

Acknowledgements

I would like to acknowledge my supervisor Professor Malcolm D. Walkinshaw for providing me with the opportunity to work on this project as well as his support and guidance. I would also like to thank the staff of the EPPF and BCF for their assistance with the protein purification and characterisation work, and in particular Sandra Bruce for taking the time to proof read this thesis (I think I owe you a few more bars of chocolate!).

In addition, I thank all members of the Walkinshaw lab, past and present, who help to maintain an incredibly knowledgeable and friendly group. In particular Iain McNae for assistance with crystallography processing.

I would also like to thank my second supervisor Professor Robert P Millar and the members of the Millar lab from QMRI Little France who were an invaluable resource during the mammalian cell culture and assay work, I give special thanks to Robin Seller for providing his time and skills in assisting with these assays.

Finally I wish to thank my family and friends for their unwavering support; I dedicate this thesis to you all.

List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
β_2AR	β_2 adrenergic receptor
BAFF	B cell-activating factor
BCMA	B Cell Maturation
BLOSUM	Blocks of amino acid substation matrix
BLyS	B lymphocyte stimulator
bp	base pairs
BSA	Bovine serum albumin
CCP4	Collaborative Computational Project Number 4
CD	Circular dichroism
cDNA	Complimentary DNA
CRD	Cys-rich domains
CSD	Cambridge Structure Database
Cys	Cysteine
DLS	Dynamic Light Scattering
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EDULISS	Edinburgh University ligand selection system
EM	Electron Microscopy
FSH	Follicle Stimulating Hormone
FP	Fluorescence Polarization
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GnRH	Gonadotropin-releasing hormone
GnRH-R	Gonadotropin-releasing hormone receptor
GPCR	G protein-coupled receptor
His	Histidine
Ile	Isoleucine
IP3	Inositol -1,4,5-trisphosphate
K_d	Dissociation constant
kDa	Kilodalton
LB	Luria-Bertani broth
IPTG	isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Leu	Leucine
LH	Luteinising hormone
LIDAEUS	Ligand Discovery at Edinburgh University
logP	octanol-water partition coefficient

MAGUK	Membrane associated guanylate kinase
Met	Methionine
NMR	Nuclear Magnetic Resonance
OD_{600}	optical cell density measured at 600nm
PBS	Phosphate Buffered Saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethyleneglycol
Phe	Phenylalanine
Pyk2	Proline rich tyrosine kinase
Pro	Proline
RMSD	Root mean square deviation
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
SH3	Src Homology 3
STP	Surface triplet propensities
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
UCSF	University of California at San Francisco
UFSR	Ultra fast shape recognition
UFSRAT	Ultra fast shape recognition with atom type
Val	Valine
V_m	Matthews coefficient

Table of Contents

Chapter 1. Utilising virtual screening to aid the discovery of small molecule ligands to disrupt protein-protein and protein-peptide interactions 1

1.1 Virtual screening 1

1.2 An overview of virtual screening strategies 1

1.2.1 Ligand based screening approaches 2

1.2.2 Identification of pharmacophores 2

1.2.3 Similarity searching 3

1.2.4 Shape similarity 4

1.2.5 Success of Ligand based screening 5

1.3 Protein based screening approaches 6

1.3.1 Docking experiments and docking optimisation 6

1.3.2 Docking solutions are highly variable and should be interpreted with care 7

1.3.3 Applying filters to compound databases is a useful method of enrichment 8

1.3.4 Filtering compounds by similarity in the post-docking stage is an efficient way of testing the most chemical space 9

1.4 Compound solubility is a key factor in compound selection 11

1.4.1 Difficulties in determining solubility. 11

1.4.2 Computational tools for estimating solubility 12

1.4.3 Current software prediction models attempt to solve the solvation problem in a number of different ways 12

1.5 Target selection for *in silico* screening 14

1.5.1 GnRH-R as a target 14

1.5.2 GnRHR binding the GnRH decapeptide controls reproductive function 16

1.5.3 Therapeutic benefits of targeting the GnRH receptor 18

1.6 Target selection for *in silico* screening - Post synaptic density protein 95 (PSD-95) 19

1.6.1 PSD-95 is a scaffolding protein which belongs to the Membrane Associated Guanylate Kinase (MAGUK) family 19

1.6.2 MAGUK proteins contain novel SH3 domain structures 20

1.6.2 PSD-95 binds to a number of therapeutically significant proteins 21

1.7 Target selection for *in silico* screening – B Lymphocyte Stimulator (BLyS) 22

1.7.1 BLyS is a member of the Tumour Necrosis Factor (TNF) family 22

1.7.2 Maturation of B cells is an important part of immunity 23

1.7.2 Targetting BLyS for treatment of autoimmune disease 24

1.8 References 25

Chapter 2. Computational and biochemical protocols 32

2.1 Computational tools and approaches 32

2.2 MODELLER 32

2.2.1 Alignment of target and template sequences in MODELLER 33

2.2.2 Calculating spacial restraints in the template structure 34
2.2.3 Generating the target model from spatial restraints of a homologous structure. 34
2.3 ProsaWeb 36
2.4 LIDAEUS 37
2.4.1 The preen stage 39
2.4.2 The pose stage 39
2.4.3 The score stage 40
2.4.4 The Sort stage 40
2.5 AutoDock vina 41
2.6 EDULISS 43
2.7 Compound clustering by Tanimoto coefficient 44
2.8 UFSRAT 48
2.9 Surface Triplet Propensities (STP) 50
2.10 Biochemical assays techniques 52
2.10.1 Culture of COS-7 cells 52
2.10.2 Transfection of COS-7 cells with P41 plasmid 52
2.10.3 Whole cell binding assay 53
2.10.4 Inositol phosphate (IP3) assay 53
2.11 References 55

Chapter 3. Building a model of the GPCR Gonadotrpin Releasing Hormone Receptor (GnRH-R). 58
3.1.1 GPCR modelling using rhodopsin as a template structure 59
3.1.2 Subsequent crystallisation of rhodopsin-like GPCR proteins offers greater insight to GPCR structure 59
3.1.3 GPCR activation isa multi-state structural event 61
3.2 Alignment of GPCR sequences shows low homology outwith key highly conserved domains 62
3.2.1 Microdomains and switch regions are conserved sequences which apply structural constraints to maintain inactive conformations 66
3.3 Alignment of GPCR crystal structures highlights favourable overlap of seven transmembrane helices. 67
3.4 Discussion of sequence and structure alignments of the GPCR proteins 68
3.5 Building a homology model of GnRH-R using a rhodopsin template 69
3.5.1 Rhodopsin and GnRH-R share low sequence homology 69
3.5.2 Homology modelling of the GnRH receptor based on rhodopsin 70
3.5.3 Evaluation of GnRH-R rhodopsin-based homology models 72
3.5.4 Flexible intracellular and extracellular loop regions present challenges in the model building process 73
3.5.4.1 Loop optimisation 74
3.5.5 Preserving ionic locks and microswitch domain interactions 74
3.6 GnRH-R modelling using the β2-adrenergic receptor 76

3.6.1 Structure generation and loop optimisation of the GnRH-R model based on the β2-adrenergic receptor ... 76
3.7 A comparison between models based on rhodopsin and the β2-adrenergic receptor ... 78
3.7.1 Superposing the ligand free homology models of GnRH-R from rhodopsin and the β2-adrenergic receptor ... 79
3.7.2 Examining differences in positioning of the conserved residues between model structures ... 79
3.8 Understanding GPCR activation and signalling ... 80
3.8.1 Insights into GnRH-R signalling from the newly crystallised β2-adrenergic receptor-Gs protein complex ... 81
3.9 References ... 85

Chapter 4. Theoretical and experimental studies on GnRH and agonist and antagonist peptide structures. ... 91
4.1 Introduction ... 91
4.2 Comparisons of GnRH receptor agonist and antagonist peptides ... 92
4.3 Generating multi-conformers of peptide antagonist/agonists. ... 97
4.3.1 Generation of Antagonist multi-conformer peptide structures ... 97
4.3.2 Generation of agonist multi-conformer peptide structures. ... 98
4.4 Comparing the agonist and antagonist multi-conformer peptide structures ... 99
4.5 Docking agonist and antagonist peptide conformers using LIDAEUS and AutoDock vina ... 100
4.5.1 LIDAEUS and AutoDock vina allow for rigid body and flexible docking environments ... 101
4.5.2 Preparing the model for docking ... 101
4.5.3 Ligand site points specify binding regions on the receptor model ... 103
4.5.4 Agonist and antagonist docking results using flexible poses in AutoDock vina using the β_2 adrenergic receptor homology model ... 104
4.5.5 Agonist and antagonist docking results using flexible poses in AutoDock vina using the rhodopsin homology model ... 108
4.5.6 Comparing Antagonist peptide rigid body docking results from LIDAEUS using rhodopsin and β_2 adrenergic receptor homology models ... 113
4.5.6.1 Antagonistic peptide docking into the Rhodopsin derived GnRH-R model ... 113
4.5.6.2 Antagonistic peptide docking – β_2 adrenergic receptor model ... 114
4.6 Crystallisation and characterisation of GnRH peptide ... 116
4.6.1 GnRH I [D-Lys6,Trp7,Leu8] peptide crystallisation strategy ... 116
4.6.2 Characterisation of GnRH peptide by circular dichroism ... 123
4.8 References ... 125

Chapter 5. Identification of small molecule inhibitors to GnRH-R using similarity searching. ... 130
5.1 Elagolix is a non-peptide orally active GnRH-R antagonist. ... 130
5.1.1 Elagolix development through structure activity relationships. ... 130
5.1.2 Docking the Elagolix molecule into the GnRH Receptor ... 131
5.1.3 Thienopyrimidinedione binding site analysis ... 132

5.1.4 Docking Elagolix using AutoDock vina 133
5.1.5 Comparing docked poses of thienopyrimidinedione scaffold and Elagolix 134
5.1.6 Probing the antagonist activity of Elagolix 134
5.2 Using the EDinburgh University LIgand Selection System (EDULISS) as a search space for compound similarity. 136
5.3 Searching EDULISS for compounds similar to Elagolix using Ultra Fast Shape Recognition with Atom Type (UFSRAT) and docking with AutoDock vina 136
5.4 Assaying compounds structurally similar to Elagolix with the GnRH receptor. 139
5.4.1 Whole cell binding assay 139
5.4.2 Inositol Phosphate inhibition (IP3) Assay 141
5.4.3 Discussion of assay results for compounds selected by Elagolix similarity 141
5.5 Quinolone compounds as potential GnRH-R antagonists 142
5.5.1 Selection of quinolone/quinolone-like compounds and AutoDock vina scoring 143
5.5.2 Assay of quinolone compounds 143
5.5.3 Discussion on the potential of quinolone compounds 145
5.6 Docking beta blockers into the "β"$_2$ adrenergic receptor derived GnRH receptor model demonstrates bias for β Adrenergic receptor ligands. 146
5.6.1 Comparison of β adrenergic receptors 146
5.6.2 Docking the B1 Selective beta blocker Bisoprolol into the GnRH receptor. 149
5.6.3 Docking the B2 Selective beta blocker Butaxamine into the GnRH receptor model. 150
5.6.4 Docking the non-Selective beta blocker Propranolol into the GnRH receptor model. 151
5.6.5 Inositol phosphate production assay indicates no GnRH receptor agonist or antagonist activity from Propranolol. 152
5.6.6 Homology models have a degree of bias for the template structure ligands 152
5.7 Conclusions and future work 154
5.8 References 156

Chapter 6. Identifying small molecule inhibitors using a LIDAEUS database mining approach 157
6.1.1 Docking the GnRH-R model derived from rhodopsin in the active conformation with EDULISS compounds using LIDAEUS 158
6.1.2 Compound selection for assay 159
6.1.3 Assay of potential inverse agonist compounds from active state rhodopsin homology model 160
6.2 Docking the GnRH-R model derived from rhodopsin in the inactive conformation with EDULISS compounds using LIDAEUS 163
6.2.1 Compound selection for assay 164
6.2.2 Assay of potential antagonist compounds from inactive rhodopsin homology model 165
6.3 Docking the GnRH-R model derived from the β_2 adrenergic receptor in the inactive conformation with EDULISS compounds using LIDAEUS 169
6.3.1 Compound selection for assay 170
6.3.2 Assay of potential antagonist compounds from β_2 adrenergic receptor homology model 173

6.4 Comparison of results from each model structure 174
6.5 References 177

Chapter 7. Biochemical analysis and structural characterisation of Post Synaptic Density protein 95 (PSD95) 177
7.1 Introduction 177
7.1.1 Canonical SH3 domains 179
7.2 Identifying a druggable construct of PSD95 182
7.2.1 Alignment of the PSD95 SH3 domain with canonical SH3 domains 184
7.3 Expression, purification and characterisation of a drug target region of PSD95 188
Construct design 188
7.4 Cloning, expression and purification of constructs A, B and C 188
7.4.1 Construct A of PSD95 188
7.4.2 Construct B of PSD95 191
7.4.3 Construct C of PSD95 194
7.4.4 Discussion of PSD95 constructs and purification methods 197
7.5 Characterisation of PSD95 construct A 197
7.5.1 Gel filtration analysis of construct A 197
7.5.2 Thermal denaturation of construct A 198
Protein concentration screening 199
DMSO concentration screening 200
7.5.3 Circular dichroism analysis of construct A 202
7.5.4 Discussion on the characterisation of construct A 203
7.6 LIDAEUS docking into the PSD95 SH3 domain 203
7.6.1 Defining the site points of PSD95 SH3-HOOK-GK 203
7.6.2 LIDAEUS results using multiple conformers and flexible docking 204
7.7 Assay development for the in vitro testing of small molecule compounds from virtual screening 208
7.7.1 Using intrinsic tryptophan fluorescence as an assay technique for LIDAEUS compounds 208
7.7.2 Using competition binding for LIDAEUS compound testing 211
7.7.2.1 Cloning KA2 C-terminal construct 212
7.7.2.3 Overexpression and purification of the KA2 C-terminal construct 213
7.7.2.4 Assessing interaction of the KA2 fragment and construct A by native gel 215
7.8 Conclusions and future work 215
7.9 References 217

Chapter 8. Biochemical and structural characterisation of B Lymphocyte Stimulating Factor (BLyS) 221
8.1 Introduction 221
8.2 Druggability of the BLyS/BAFF-R complex 222

8.3 Expression, purification and characterisation of BLyS 224
8.3.1. Expression 224
8.3.2 Refolding and purification 225
Protein refolding by dilution 225
On column refolding 226
8.3.3 Characterisation of the BLyS 60-mer complex 227
8.3.3.1 Analytical gel filtration 227
8.3.3.2 Thermal denaturation 229
8.3.3.3 Intrinsic fluorescence spectroscopy 231
8.3.3.4 Comparisons of emission spectra between refolded and denatured BLyS 231
8.3.3.5 Transmission electron microscopy (TEM) analysis 233
8.3.3.6 Dynamic light scattering 234
8.4 Virtual screening of the BLyS structure with LIDAEUS 236
8.4.1 Defining the site points for BLyS 236
8.4.2 Analysis of the Leu28 binding pocket reveals a suitable target for docking 236
8.4.3 LIDAEUS docking of the EDULISS database to the BLyS protein 238
8.5 Ligand selection for BLyS/BAFF-R disruption screening 242
8.6 Measuring the effect of virtual screening compounds on BLyS fluorescence 244
8.7 Conclusions and future work 247
8.8 References 249

List of Figures

Figure 1-1. Pharmacaphore searching focuses on the positional space that particular chemical groups occupy. 3
Figure 1-2. Fingerprint generation in similarity searching is a method of describing atom types. 4
Figure 1-3. A simplified example of atom pairing for similarity searching. 5
Figure 1-4. A typical cluster dendrogram of a virtual screening docking experiment. 10
Figure 1-5. The human type I GnRH Receptor. 16
Figure 1-6. Sex steroid production is controlled by GnRH binding the cognate receptor. 17
Figure 1-7. Sequence variation in GnRH decapeptides. 18
Figure 1-8. MAGUK scaffolding within the postsynaptic density. 20
Figure 1-9. Schematic diagram comparing the beta stand orientation in the canonical SH3 domain (a) and in the MAGUK protein (b). 21
Figure 1-10. Intertwined structure of the SH3 domain in PSD-95. 21
Figure 1-11. BLyS association with the cognate receptor and the TACI and BCMA receptors. 23
Figure 2-1. An overview of the model structure building process in the MODELLER program. 33
Figure 2-2. LIDAEUS docking pipeline. 37
Figure 2-3. Site-points created for the rhodopsin based GnRH receptor model. 39
Figure 2-4. Defining rotatable bonds of a sample ligand structure using AutoDockTools. 41
Figure 2-5. Defining docking area of the target protein using AutoDockTools. 42
Figure 2-6. Summary of the drug likeness criteria met by the compounds in the EDULISS database. 44
Figure 2-7. A clustering of 10 compounds by Tanimoto coefficient. 48
Figure 2-8. Similarity searching the EDULISS database using UFSRAT and the Elagolix structure.. 50

Figure 2-9. Scoring of the surface of the binding pocket of GnRH-R with STP. 51
Figure 3-1. Overall topology of the GPCR rhodopsin family. 58
Figure 3-2. Structural conformation changes in the activation of the β2 Adrenergic receptor 62
Figure 3-3. Sequence homology of the six GPCR crystal structures. 64
Figure 3-4. Analysis of ionic locks and microdomains in GPCR proteins. 67
Figure 3-5. Structural alignment of GPCRs using rhodopsin as the anchored protein. 68
Figure 3-6. Identity scoring of GnRH-R and rhodopsin residues. 70
Figure 3-7. Superposing the five GnRH-R model structures generated by MODELLER. 71
Figure 3-8. Using ProSA Web as a tool to check model quality. 72
Figure 3-9. Evaluation of binding residue placement in MODELLER generated structure from rhodopsin. 73
Figure 3-10. Comparison of microdomains between rhodopsin derived GnRH-R model and rhodopsin. 75
Figure 3-11. Alignment of GnRH-R and the β2-adrenergic receptor. 76
Figure 3-12. ProSA-Web analysis of the GnRH-R β2-adrenergic receptor homology model. 77
Figure 3-13. Structural alignment of GnRH-R models created using rhodopsin and the β2-adrenergic receptor. 78
Figure 3-14. Comparison of key conserved residues between the rhodopsin (green) and the β2-adrenergic receptor (cyan) derived homology models. 79
Figure 3-15. Comparison of Gs protein binding between β2-adrenergic receptor and GnRH-R β2-adrenergic receptor homology model. 83
Figure 4-1. Comparison of 4 antagonist peptide structures with GnRH-I shows that a consensus of 3 key amino acids are required for binding to the receptor whilst avoiding activation of signalling. 93
Figure 4-2. 2D structures of clinically approved GnRH-R antagonist peptides. 94
Figure 4-3. Structural comparison of four high potency GnRH analogue agonist peptides. 96
Figure 4-4. GnRH-I residue map and how residue substitution affects binding. 97
Figure 4-5. Superimposed conformer structures for each of the antagonist peptide. 98
Figure 4-6. Superimposed conformer structures for each of the agonist peptides. 99
Figure 4-7. Comparison of energy minimised conformation of four antagonist and four agonist peptides against GnRH. 100
Figure 4-8. GnRH-R model structure used for docking. 102
Figure 4-9. Site-points created for the GnRH receptor. 103
Figure 4-10. Overlay of four agonist peptides in their respective docked conformations using AutoDock vina. 104
Figure 4-11. Overlay of four antagonist peptides in their respective docked conformations using AutoDock vina. 105
Figure 4-12. Potential interactions between docked agonist peptides and the gonadotrophin releasing hormone receptor using a β2 adrenergic receptor derived model. 106
Figure 4-13. Potential interactions between docked antagonist peptides and the gonadotrophin releasing hormone receptor using a β2 adrenergic receptor derived model. 107
Figure 4-14. Identifying hotspots in agonist and antagonist peptide binding to the GnRH-R β2 adrenergic receptor homology model. 107
Figure 4-15. Overlay of four agonist peptides in their respective docked conformations using AutoDock vina. 109
Figure 4-16. Overlay of four antagonist peptides in their respective docked conformations using AutoDock vina. 109
Figure 4-17. Potential interactions between docked agonist peptides and the gonadotrophin releasing hormone receptor using a rhodopsin derived model. 110
Figure 4-18. Potential interactions between docked antagonist peptides and the gonadotrophin releasing hormone receptor using a rhodopsin derived model. 111
Figure 4-19. Identifying hotspots in agonist and antagonist peptide binding to the GnRH-R rhodopsin homology model. 112
Figure 4-20. Preliminary ammonium sulphate screen to determine precipitation point. 118
Figure 4-21. Needle clusters of GnRH I [D-Lys6,Trp7,Leu8] peptide in 0.2 M Zinc acetate dihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 18% w/v Polyethylene glycol 8,000 at 4 °C. 119
Figure 4-22. Diffraction image of GnRH I [D-Lys6,Trp7,Leu8] peptide needles. 120
Figure 4-23. 2 peptide structures with space group P212121 were found in the CSD. 122

Figure 4-24. Circular dichroism spectra of GnRH I [D-Lys6,Trp7,Leu8] in water and TFE solution. 123
Figure 5-1. 2D structure of the orally active non-peptide antagonist Elagolix and the thienopyrimidinedione scaffold that it was built on. 131
Figure 5-2. The uracil pharmacaphore used in Elagolix has been designed for high affinity. 131
Figure 5-3. Comparison of the suggested thienopyrimidinedione scaffold binding in the β2AR derived receptor model using AutoDock vina against the literature docking study from Betz et al. 132
Figure 5-4. Docking Elagolix into the β2 Adrenergic receptor derived model using AutoDock vina shows that as in the scaffold docking experiment the elagolix molecule is positioned into the deeper pocket closer to Q208. 133
Figure 5-5. Comparison of docked binding poses between thienopyrimidinedione scaffold from this study and the literature and Elagolix in the GnRH-R active site. 134
Figure 5-6. Comparison of the docked poses of the GnRH-I peptide and Elagolix it can be seen that Elagolix occupies almost half of the space of the GnRH-I peptide. 135
Figure 5-7. Comparisons of binding locations between compounds found from EDULISS with high similarity to Elagolix with Elagolix. 138
Figure 5-8. Results of whole cell binding assay of elagolix similar compounds at 0.01, 0.1, 1.0, 30 and 250μM concentrations. 140
Figure 5-9. Radioactivity from deuterated inositol produced by stimulation with GnRH peptide. 141
Figure 5-10. Single point assay results for quinolone/quinolone like compounds. 143
Figure 5-11. Docked poses of compounds D10 and G10. 145
Figure 5-12. Sequence alignment of beta adrenergic receptors 1-3 indicates most homology is found in the membrane-bound regions of the receptor structures. 147
Figure 5-13. Superimposing the Turkey β1 adrenergic receptor (PDB code 2VT4) and the Human β2 adrenergic receptor shows an RMSD value of 0.809Å. 148
Figure 5-14. Docking the Bisoprolol molecule into the GnRH receptor indicated that one hydrogen bond could be formed between the ligand and receptor. 150
Figure 5-15. Docking the blocker Butaxamine into the GnRH receptor suggests hydrogen bonding to Asn212. 150
Figure 5-16. Docking the unselective beta blocker Propranolol into the GnRH receptor suggests hydrogen bonding to Leu286. 151
Figure 5-17. Comparison of Inositol phosphate production between the basal state and the 250μM Propranolol incubated cells shows that IP production is not significantly affected. 152
Figure 5-18. Comparison of docked beta blocker positions and scores between GnRH-R models derived from the β2-adrenergic receptor and rhodopsin shows higher scoring in the β2-adrenergic receptor model. 153
Figure 6-1. Positions of the key GnRH binding residues in each of the three model structures. 157
Figure 6-2. Structure of the active conformation rhodopsin derived GnRH receptor model used in this part of the docking study. 158
Figure 6-3. Clustering of active state GnRH-R from homology modelling with rhodopsin. 159
Figure 6-4. Results of IP assay for 100 μM antagonist compounds from rhodopsin active state model. 161
Figure 6-5. Radioligand binding assay of compounds produced from EDULISS screening of the GnRH-R rhodopsin model. 162
Figure 6-6. Clustering of inactive state GnRH-R from homology modelling with rhodopsin. 164
Figure 6-7. Binding assay of compounds derived from rhodopsin homology model. 165
Figure 6-8. IP inhibition results for antagonist compounds derived from the inactive GnRH-R rhodopsin homology model. 167
Figure 6-9. Binding assay of compounds derived from rhodopsin homology model. 168
Figure 6-10. Structure of the inactive GnRH receptor model derived from β_2AR. 169
Figure 6-11. Clustering of inactive state GnRH-R from homology modelling with β_2AR. 170
Figure 6-12. One point assay results of compounds selected using the β_2AR model structure. 174
Figure 6-13. Comparison of docked positions of the most active compounds. 175
Figure 7-1. A proline rich region of PI(3)KC2α is able to associate to the SH3 domain of PSD95 and control inflammatory response. 179
Figure 7-2. SH3 domain from ABL tyrosine kinase (PDB ID 1ABO). 180
Figure 7-3. Surface view of SEM5 protein with residues which undergo chemical shift on ligand binding shown as sticks. 181

Figure 7-4. Structure of the SH3 domain small molecule inhibitor UCS15A. 182
Figure 7-5. Binding propensity of the SH3-Hook-GK region of PSD95 as coloured by STP. 183
Figure 7-6. Comparison of binding orientations of class I and II SH3 domains to PSD-95. 183
Figure 7-7. Sequence alignment of canonical SH3 domains (both Class I and II) and PSD-95 SH3 domain. 185
Figure 7-8. Overlay of peptides structures which bind to SH3 domains aligned in Figure 7-7. 186
Figure 7-9. Amino acid distribution of the three residues before and after the PxxP motif binding peptides. 187
Figure 7-10. Amino acid distribution within the PxxP motif. 187
Figure 7-11. Construct design of druggable PSD-95 recombinant protein. 189
Figure 7-12. Cloning of SH3-Hook-GK construct. 190
Figure 7-13. Purification of SH3-Hook-GK fragment of PSD95. 191
Figure 7-14. PCR product of SH3 domain alone was approximately 250bp in length. 192
Figure 7-15. . Expression trials with construct B showed highest expressions from C41(DE3) cells induced for 4 hours at 37°C.. 192
Figure 7-16. Refolding and purification profile of SH3 domain construct B.. 194
Figure 7-17. Whole cell pellet SDS PAGE gel of construct C overexpression trials. 195
Figure 7-18. Purification trace of SH3-GST and SDS-PAGE analysis.. 196
Figure 7-19. . Gel filtration analysis of construct A protein reveals the protein is primarily in a monomeric state. 198
Figure 7-20. Protein concentration effects on thermal melting. 200
Figure 7-21. Effects of increasing DMSO concentration on protein stability in thermal denaturation studies. 201
Figure 7-22. Analysis by circular dichroism shows that the recombinant protein is well folded. 202
Figure 7-23. Determination of site points for LIDAEUS docking experiments. 204
Figure 7-24. Compound clustering results for PSD-95 from LIDAEUS. 205
Figure 7-25. Potential binding interactions of 6 LIDAEUS conformers docked into the SH3 domain of PSD95. 207
Figure 7-26. The SH3 domain of PSD95 contains a tryptophan residue in the specificity pocket which is exposed to the surface. 208
Figure 7-27. Intrinsic fluorescence from LIDAEUS compounds selected for assay. 209
Figure 7-28. Effects of compound incubation on tryptophan fluorescence. 210
Figure 7-29. Cloning of SH3/GST domain construct. 213
Figure 7-30. Optimising expression of KA2 N-terminal fragment. 213
Figure 7-31. Purification of KA2 N-terminal fragment. 214
Figure 7-32. Assessment of Construct A binding to KA2 subunit using a native gel. 215
Figure 8-1. The cysteine rich domain (CRD) of the BAFF receptor. 222
Figure 8-2. The TNF homology domain of the BLyS monomer (left). The extended DE loop allows interactions between two BLyS trimers in a “handshaking” fashion. 223
Figure 8-3. Structure of BLyS trimer complexed with BAFF-R fragments. 223
Figure 8-4. Overexpression trials of BLyS protein 224
Figure 8-5. Overexpression trials for BLyS protein showed protein is still insoluble. 225
Figure 8-6. Purification results for on column refolding of BLyS from inclusion bodies. 227
Figure 8-7. Comparison of refolding state of BLyS by gel filtration analysis. 228
Figure 8-8. Silver stain results of refolded BLyS in comparison to soluble BLyS. 229
Figure 8-9. Melting profile of refolded BLyS protein with varying pH. 230
Figure 8-10. Hydrophobic profile of BLyS 60-mer. 230
Figure 8-11. Fluorescence profiles of urea denatured and refolded BLyS. 232
Figure 8-12. Transmission electron microscope images of BLyS concentration range. 234
Figure 8-13. Dynamic Light Scattering (DLS) results show two main peaks in solution. 235
Figure 8-14. Defining site points on the BLyS/BAFF-R interface. 236
Figure 8-15. Binding pocket predictions from PocketDepth analysis on BLyS. 237
Figure 8-16. Predicted binding mode between 25SPH1-178-375 and BLyS shows a number of hydrogen bonding opportunities both deep inside the pocket as well as the upper lip. 240
Figure 8-17. Predicted binding mode between 26SPH1-825-056 and BLyS shows a number of hydrogen bonding opportunities both deep inside the pocket as well as the upper lip. 241
Figure 8-18. Predicted binding mode between 35SPH1-122-641 and the BLyS protein. 8 Hydrogen bonds are possible between the compound and the protein. 241

Figure 8-19. All six compounds chosen for BLyS fluorescence study have very little fluorescent properties. 245
Figure 8-20. Effects of compounds from virtual screening on BLyS protein fluorescence. 246

List of Tables

Table 1-1. A selection of commercial and academic docking software available for in silico screening. 6
Table 1-2. Summary of error values for a collection of solubility prediction methods. 13
Table 2-1. Common similarity coefficients used in similarity searching. 46
Table 3-1. Summary of current GPCR crystal structures published in the literature. 60
Table 3-2. Pairwise alignment scores of GPCR crystal structures. 64
Table 3-3. Potential intermolecular interactions between GnRH-R and Gs. 82
Table 4-1. Properties of the four selected high potency GnRH-R antagonist peptides for docking. 92
Table 4-2. Properties of the four selected high potency GnRH-R agonist peptides for docking. 95
Table 4-3. Scoring 50 peptide conformers for the Abarelix, Antide, Cetrorelix and Ganirelix antagonists using the rhodopsin derived GnRH-R model. 114
Table 4-4. Scoring 50 peptide conformers for the Abarelix, Antide, Cetrorelix and Ganirelix antagonists using a structure modelled from the β2 adrenergic receptor. 115
Table 4-5. Crystallisation parameters for low molecular weight peptide fragments. 116
Table 4-6. Conditions which show microcrystal formation of the GnRH peptide. 118
Table 4-7. Processing statistics for GnRH I [D-Lys6, Trp7, Leu8] Peptide. 120
Table 4-8. Calculating solvent content of the asymmetric unit. 121
Table 4-9. Cell dimension parameters for the P212121 peptides from the CSD. 122
Table 5-1. The structure of Elagolix was run through UFSRAT and the top 10 similar compounds were scored in AutoDock vina for predicted affinity. 137
Table 5-2. IC50 values and structures for quinolone GnRH-R inhibitors found in the literature. 142
Table 5-3. Selected quinolone/quinolone like compounds assayed for antagonist activity. 144
Table 5-4. Identity comparisons between the Beta Adrenergic receptor sequences. 147
Table 5-5. Beta blocker compounds used to assess bias in the GnRH-R model structure. 149
Table 6-1. Compounds selected for assay from the active state structure search for antagonist ligands. 160
Table 6-2. Compounds selected for assay from the ground state structure search for inverse agonists. 166
Table 6-3. Compounds selected for assay from LIDAEUS β2AR GnRH-R model. 171
Table 6-4. Affinity scoring for each model compound which produced the most inhibition. 176
Table 7-1. Identity scores between PSD95 SH3 and 7 other class 1 and class 2 SH3 domain proteins. 185
Table 7-2. Selected Chembridge compounds to be assayed against PSD95. 206
Table 8-1. Structures and scores for the top 10 scoring compounds from BLyS/BAFF-R binding pocket docking. 238
Table 8-2. List of compounds selected for assay with BLyS protein. 242

Chapter 1. Utilising virtual screening to aid the discovery of small molecule ligands to disrupt protein-protein and protein-peptide interactions

1.1 Virtual screening

The average cost of developing a new drug from the disease targeting stage through to becoming a marketable product is over $800m[1] and takes around 14 years[2]. A considerable focus of pharmaceutical companies is therefore to endeavour to reduce costs. One of the large costs associated with creating a new drug is in the number of failures which are encountered in the research workflow. Because of the demand to reduce the failure rate of a drug there is pressure to increase the potential success rate before the drug molecule itself is even synthesised. One method of doing this is by implementing a virtual screening strategy into the process. This has been quite successful and has led to the discovery of a number of chemically active molecules for key target proteins[3].

1.2 An overview of virtual screening strategies

Whilst there is overlap in the overall result of virtual screening strategies the detailed methods of ligand selection and compound enrichment can be quite different. The different approaches may be tailored to suit the nature of the target protein depending on how well the protein has been documented.

1.2.1 Ligand based screening approaches

In many cases the crystal structure of the target protein of interest may not be available. This is particularly true for membrane proteins which make up a considerable proportion of drug targets[4], perhaps because purification as a recombinant protein is difficult or because the correct conditions have not yet been empirically derived. A ligand based screen is a useful approach for such proteins, however lack of receptor information is not the only reason to use this approach. If there are already well documented chemically active ligand molecules then the idea of similar molecule having similar chemical properties[5] to generate new novel compounds can be exploited.

1.2.2 Identification of pharmacophores

The idea of querying structure databases by pharmacophore, a particular arrangement of chemical groups (Figure 1-1), is not a new concept[6]. Since 1985 there have been publications which have described the use of computational methods to derive pharmacophores for serotonin inhibitors in the development of antidepressants[7] and beta blockers[8]. However it is becoming a more useful means of identifying potential compounds now that large chemical databases are easier to access. The recently published Pharmer program[9] which can search the Zinc database[10] of 2.7 million commercially available compounds is a step in the right direction. Typically at least one known active compound is needed in order to generate a pharmacophore fingerprint. However further success is likely if additional compounds are known so that a more detailed picture of the receptor pharmacophore can be drawn.

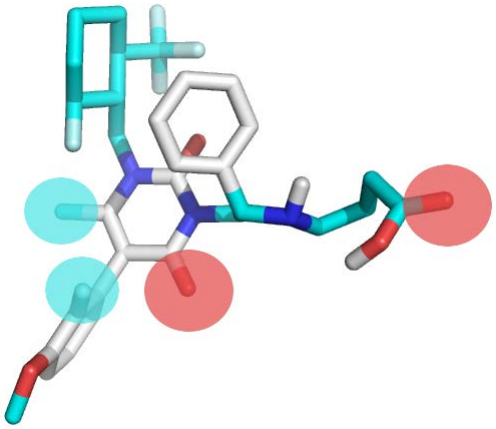


Figure 1-1. Pharmacaphore searching focuses on the positional space that particular chemical groups occupy in order to interact with the target molecule, disregarding the other properties of the molecule. In this example the chemical groups circled in blue and red are important for activity, their 3D positions are used as search parameters to find molecules with similar chemical groups that occupy these positions.

1.2.3 Similarity searching

In order to search for a molecule which is similar to a known active a way has to be found to describe the active compound in order to search the database. This can then be compared to another string using for example a Tanimoto Coefficient to determine similarity[11]. A common way of doing this is to use "fingerprints", reading in a two dimensional structure and obtaining an output text string descriptor as described in Figure 1-2. The Daylight fingerprint algorithm is an example of this. However a number of fingerprinting algorithms exist which generate the text string in a number of different ways, the efficacy of which are dependent on the user's requirements[12].

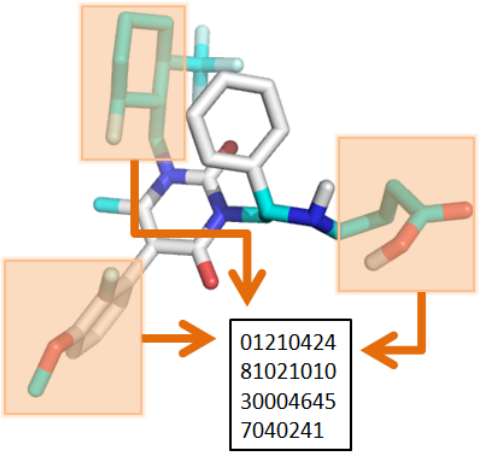


Figure 1-2. Fingerprint generation in similarity searching is a method of describing atom types. The fingerprint generated is a textual interpretation of the molecule which can be compared against other compound fingerprint text strings to measure similarity.

One of the disadvantages of this type of sorting was the time taken to generate fingerprints.

1.2.4 Shape similarity

More recently a new technique has been developed called Ultrafast Shape Recognition[13] which is a more rapid method of shape comparison because it does not rely on superposition of the starting ligand against the structures in a target database, a factor which not only reduces computational time but also removes the possibility for error in superposition[14]. Instead of considering the molecule as a solid 3D shape it breaks it down into a series of bound atoms (Figure 1-3). The main benefit of this technique is derived from the way that atom distances are measured between four key locations; the "molecular centroid (ctd), the closest atom to ctd (cst), the farthest atom to ctd (fct), and the farthest atom to fct (ftf)"[15]. This allows the maximum distances from the centre of the molecule to be calculated for rapid comparison.

This method has proven to be very successful in quickly analysing large datasets for similar structures in comparison to the traditional superposing approach[13].

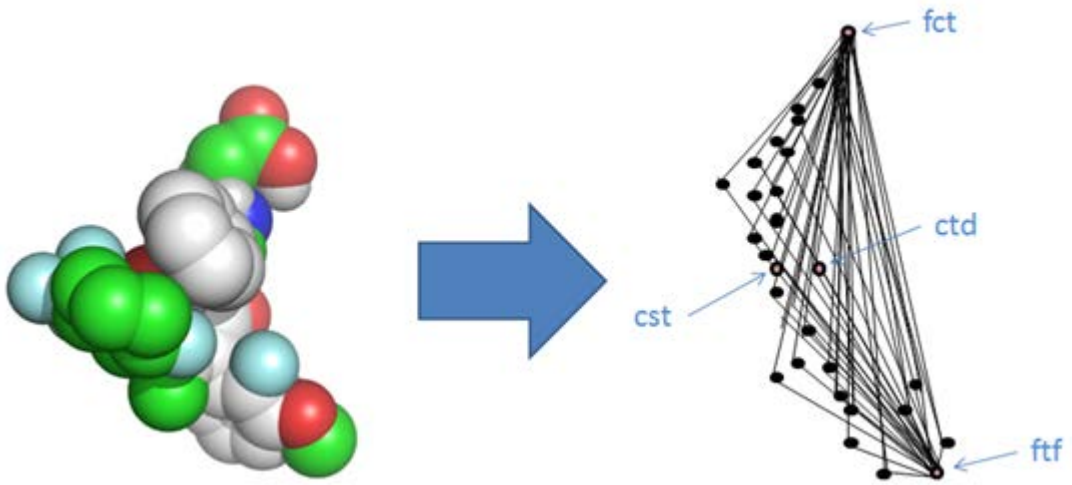


Figure 1-3. A simplified example of atom pairing for similarity searching. In a typical example all atom pair distances are calculated with respect to four reference locations: the molecular centroid (ctd), the closest atom to ctd (cst), the farthest atom to ctd (fct) and the farthest atom to fct (ftf) thus eliminating the need for superposition regardless of ligand size[15].

1.2.5 Success of Ligand based screening

Although as previously discussed the ligand based approach was slightly less favourable than being able to screen against an x-ray crystal structure of a receptor there have still been a number of reported active molecules found using this approach. Recent examples with possible therapeutic benefits include the discovery of novel IKK-β inhibitor which has been implicated in the management of inflammation and cancer through control of NF-κβ activation[16]. Another example of a cancer target derived from ligand based screening is that of inhibitors of human tyrosyl-DNA phospodiesterase (hTdp1)[17]. In this case pharmacophore filtering was used with fourteen known inhibitors to increase the likelihood that the compounds found would fit the binding profile for the active site.

1.3 Protein based screening approaches

1.3.1 Docking experiments and docking optimisation

If a crystal structure of the target protein is available then it is possible to go straight into docking optimisation and binding experiments. One interesting problem of molecular docking with respect to small ligands and their large protein targets is the number of different docking programs which exist (Table 1-1) to attempt to address the problem of predicting binding affinity. Reviews have drawn comparisons with regard to the efficiency and accuracy of a number of these programs[18, 19]. There have been noted difficulties in making a valid comparison[20], commonly the use of root mean square deviation between crystal structure ligand sites and the predicted pose from docking is used to assess docking success. However when dealing with small molecule ligands the changes in root mean square deviation can be small even if the ligand molecule is in a completely different orientation in the docked pose to the crystal structure.

Table 1-1. A selection of commercial and academic docking software available for *in silico* screening.

Program	Affiliation	Website
Autodock	The Scripps Institute	http://autodock.scripps.edu/
DOCK	UCSF	http://dock.compbio.ucsf.edu/
FlexX	BioSolveIT	http://www.biosolveit.de/flexx/

Fred	OpenEye Scientific Software	http://www.eyesopen.com/fred
Glide	Schrödinger	http://www.schrodinger.com/products/14/5/
GOLD	Cambridge Crystallographic Data Centre	http://www.ccdc.cam.ac.uk/products/life_sciences/gold
ICM	MolSoft	http://www.molsoft.com/docking.html
LIDAEUS	University of Edinburgh	http://opus.bch.ed.ac.uk/lidaeus
MSU-SLIDE	Michigan State University	http://www.bch.msu.edu/~kuhn/software/slide
Surflex-Dock	Tripos	www.tripos.com

1.3.2 Docking solutions are highly variable and should be interpreted with care

One of the more frequently encountered difficulties when using molecular docking software is the lack of consistency of the results across a range of different programs. Because there is such a variety of ways to score molecular interactions the docking results are frequently open to interpretation. Should more emphasis be placed on the results obtained from commercial vendors rather than freely available software?

Comparison between the commercial FlexX against Glide and Dock, which are available as academic licenses, indicates this is not the case[21]

This presents the problem of results being biased towards the scoring system being used[22]. One way of getting around this problem is by using multiple docking and scoring programs in the virtual screening stage. Whilst this is more computationally expensive it has the advantage of providing a more thorough validation of target docking results which should maximise the likelihood of poorly active compounds being removed in the "fail early" stage[3].

1.3.3 Applying filters to compound databases is a useful method of enrichment

A good docking score is not the only important factor in selecting compounds to take forward for testing. For example they may possess undesirable reactive chemical groups (such as an epoxide ring) or their molecular properties may pose difficulties in crossing the blood-brain barrier. It may also be advantageous to check for compounds which may interfere with assay results (for example they may possess delocalised ring structures which could affect a fluorescence assay). In order to reduce the unnecessary inclusion of these types of compounds, which can account for more than half of a commercial database[23], it is often desirable to screen the database to remove any entries which do not meet the drug likeness criteria.

A number of screens have been developed to make this process easier. The most widely known is the Lipinski rule of five which, since publication in 1997 has been cited over four thousand times[24]. The foundation of this rule is that drug like compounds are more likely if they have less than 5 H-bond donors, no more than 10

H-bond acceptors, molecular weight is below 500 daltons and the octanol-water partition coefficient (LogP) value should not exceed 5. However since these rules were published there have been a number of additional studies on the general properties of drug like compounds. Ghose et al[25] and Oprea [26] further tuned into the drug like space in chemical databases by extending the rules to include optimum values for rotatable bonds, molar refractivity and total atom number. Fortunately software such as Instant JChem (Chemaxon) or the FAF-Drugs2 server[27] exist which can do this process efficiently, being able to combine multiple drug likeness criteria as well as recognition of undesirable functional groups[28].

It should be noted however that overly stringent drug likeness filters will reduce the finding of more novel scaffolds which can be adapted at a later time to increase bioavailability. It can therefore be advantageous to examine compounds which border the boundaries of the database manually for their potential. Obviously this process is more suited to smaller databases where the number of compounds which may need to be manually examined are manageable. However this method also introduces a variable of human bias. A means of avoiding this is to employ a "machine learning" strategy where a software program is designed to look at a series of drug like and non-drug like compounds and create a set of rules to apply. While this method is still in relative infancy it is finding uses in drug discovery, most recently in work involving Hepatitis C[29].

1.3.4 Filtering compounds by similarity in the post-docking stage is an efficient way of testing the most chemical space

After a list of the top hits from docking studies is established it is common to assess how similar each compound is in respect to one another. Whilst it would be ideal to

test all top hits in an assay this is both financially and logistically challenging. Instead the compounds can be clustered based on similarity. If similar compounds share similar properties, testing all of the compounds from a highly similar cluster is unnecessary. In practice there are some exceptions to this paradigm, for example the addition of a methyl group to a compound may appear to be a small change but when added to fluorosulfonate to form methyl fluorosulfonate (magic methyl) the change in activity is huge.

Figure 1-4 shows an example of clustering using fingerprinting and then organising by similarity using a Tanimoto coefficient.

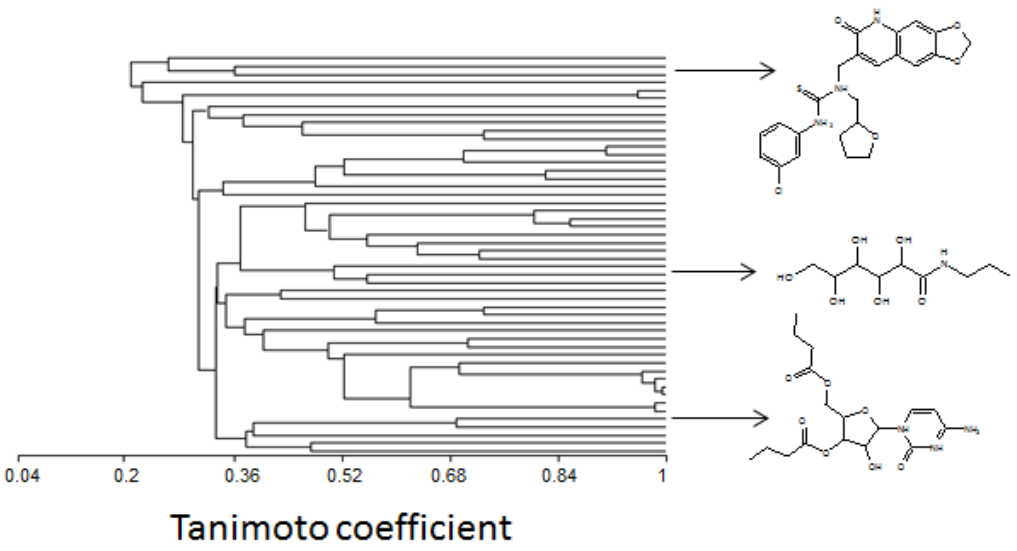


Figure 1-4. A typical cluster dendrogram of a virtual screening docking experiment. This example shows the distribution of similarity by using the daylight fingerprint for each compound and then scoring similarity according to a tanimoto coefficient, therefore highly similar compounds score closer to 1.

1.4 Compound solubility is a key factor in compound selection

Whilst an assortment of tools are available which allow rapid elimination of undesirable compounds based on the likelihood of reacting to form covalent bonds and other errant behaviour, the accurate prediction of water solubility is still elusive.

1.4.1 Difficulties in determining solubility.

The most accurate way of determining solubility is by measuring it directly. This is difficult however when only a small quantity of (possibly expensive) sample may be available. The process itself is laborious and time consuming[30], particularly if the sample must be reclaimed from the solvent.

If the general solubility equation (Equation 1) is used solubility can be estimated from the melting point and the LogP of the compound. This is quicker than direct measurement since the LogP is often given by the supplier and offers reasonable accuracy[31]. However the sample will likely be irreversibly altered during the melting process and this is not necessarily the most applicable equation for small molecule compounds[32].

$$\log S_w = 0.5 - 0.01(MP - 25) - \log K_{ow} \qquad \text{(Equation 1)}$$

Where S_w is the molar aqueous solubility, MP is the Melting Point (Celsius) and K_{ow} is the octanol–water partition coefficient (LogP).

1.4.2 Computational tools for estimating solubility

The ability to accurately predict solubility from using structural information is important in the virtual screening process as poor solubility of samples is a major practical problem in screening for binding.

In the development of a predictive model a chemically diverse training set is required in order to gain the best structure activity hypothesis. Unfortunately the solubility information of compounds is often held confidentially by the supplier or the data just does not exist. Recently there have been joint ventures between companies such as Specs and Pharma Algorithms to launch a DMSO solubility predictor[33]. The manufacturer claims an overall prediction accuracy of 82% although the software is commercially licensed and has not yet been subject to review in the literature.

1.4.3 Current software prediction models attempt to solve the solvation problem in a number of different ways

A recent review of solubility prediction programs provides a survey showing how different groups have attempted to model the compound/solvent interaction[34]. Each group developed a prediction model using a large variation in training set size (the smallest training set size was 150 while the largest was 2874). After their model was generated it was used to predict the solubility of 21 compounds frequently used to test solubility models including sparingly soluble compounds such as benzocaine and DDT, fairly soluble compounds such as aspirin and highly soluble compounds such as antipyrine. This gave a solubility range of 2.45 mol/ L to 0.008 µM/ L. In each case the log units of molar solubility (logS) were calculated and compared to the experimentally derived logS values and the standard error was found. (Table 1-2).

Table 1-2. Summary of error values for a collection of solubility prediction methods. Figure modified from [34].

Lead author	Descriptors used	Training set size	Test set size	Standard error (log)
Huuskonen	2D topological	884	413	0.6
Hou	Atomic	1290	120	0.79
Jorgensen	Whole molecule	150	149	0.72
Yan	3D descriptors	797	496	0.59
Wegner	2D topological	1016	253	0.54
Delaney	Whole molecule	2874	528	0.96
Klopman	2D substructural	1168	120	0.79
Liu	2D topological	1312	258	0.72
Butina	Whole molecule	2688	640	1.01
Klamt	Quantum mechanical	150	107	0.61

The range in standard error is 0.54-1.01 whilst the range in training set size varies from 150-2874. Given how small some of the training sets are this low error is fairly

encouraging, especially if we consider that experimentally derived solubility data can vary by as much as 1 log unit[35]. More interesting however is how the three lowest standard errors were seen in models which were all created using different approaches.

1.5 Target selection for *in silico* screening

The overall aim of this study is to identify small molecule inhibitors to three proteins. The first target is the Gonadotropin releasing hormone receptor which is a membrane bound protein controlled by a receptor-peptide binding event which regulates reproductive function[36] and has been implicated in the treatment of prostate cancer[37].

The next target is postsynaptic density protein 95 (PSD-95). This is a commonly found scaffolding protein found near the membrane of postsynaptic neurons and has a novel SH3 domain which has been shown to mediate inflammatory pain via protein-protein interactions[38].

Finally the B-lymphocye stimulating factor (BLyS) which mediates the proliferation and differentiation of of B cells[39] is found to be upregulated in a number of autoimmune diseases including systemic lupus erythematosus[40].

All three proteins are important medical targets with a need for development of small molecule inhibitors. They represent challenging interactions for disruption that virtual screening techniques can help to address.

1.5.1 GnRH-R as a target

GnRH-R is a G-protein coupled receptor of the Class A (rhodopsin-like) family. The receptor has been shown to exist as three homologs in a number of species (GnRH-R I-III) and binds to the complimentary GnRH decapeptides (GnRH I-III). Only the type one receptor is active in humans[41]. Whilst the type two receptor is present in the human genome it contains a frame shift and premature stop codon which prevents transcription of a receptor which binds the GnRH-II peptide[42]. The GnRH-II peptide has however been shown to play a role in the control of inositol-1,4,5-trisphosphate by binding to the type one receptor[43]. The type three receptor is found exclusively in fish and amphibian species[44].

GnRH-R I is a 328 amino acid protein (Figure 1-5) which unlike other members of the GPCR family does not possess a C-terminal tail. It does however possess an N-terminal region which is thought to be largely disordered. The absence of a C-terminal tail is important from a signalling perspective because this feature is often involved in the internalisation and subsequent desensitisation response. Subsequently the type I receptor displays delayed desensitisation when continually stimulated by an agonist molecule[45]. Studies have shown that the type II GnRH peptide and the cognate receptor show high conservation[46], so it can be considered that the loss of the C-terminal tail in the type I receptor was functionally important especially if it is noted that the point at which the C-terminus terminates, the YFSL motif, is known to bind to PDZ domains[47].

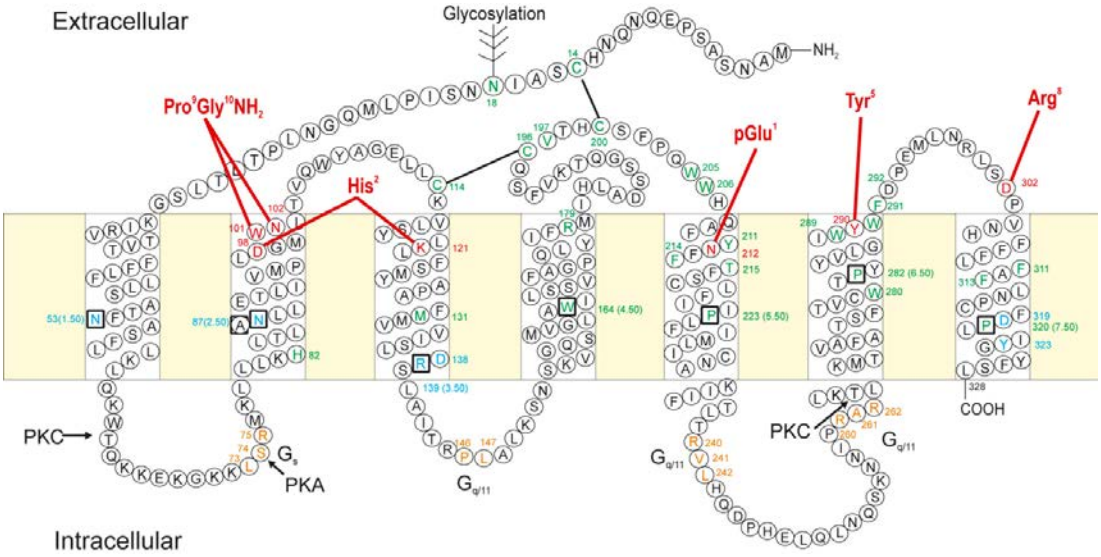


Figure 1-5. The human type I GnRH Receptor. Ligand binding sites are indicated in red, structurally significant residues are highlighted green, GPCR conserved residues are shown in orange and signalling residues are shown in blue. Diagram taken from reference [48].

1.5.2 GnRHR binding the GnRH decapeptide controls reproductive function

The 10-mer GnRH hormonal peptide is the key regulator of the reproductive system and stimulates the release of the Gonadotrophins; these are Luteinising Hormone (LH) and Follicle Stimulating Hormone (FSH)[49]. The hypophysiotropic peptide (GnRH I) has the amino acid sequence pGlu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH2 and is synthesised as a 92 amino acid pre-pro-hormone in the septal-preoptic-hypothalamic regions[50]. It is secreted in a pulsitile fashion from the median eminence approximately every 30-120 minutes and travels via the hypophyseal portal system to the anterior pituitary (Figure 1-6).

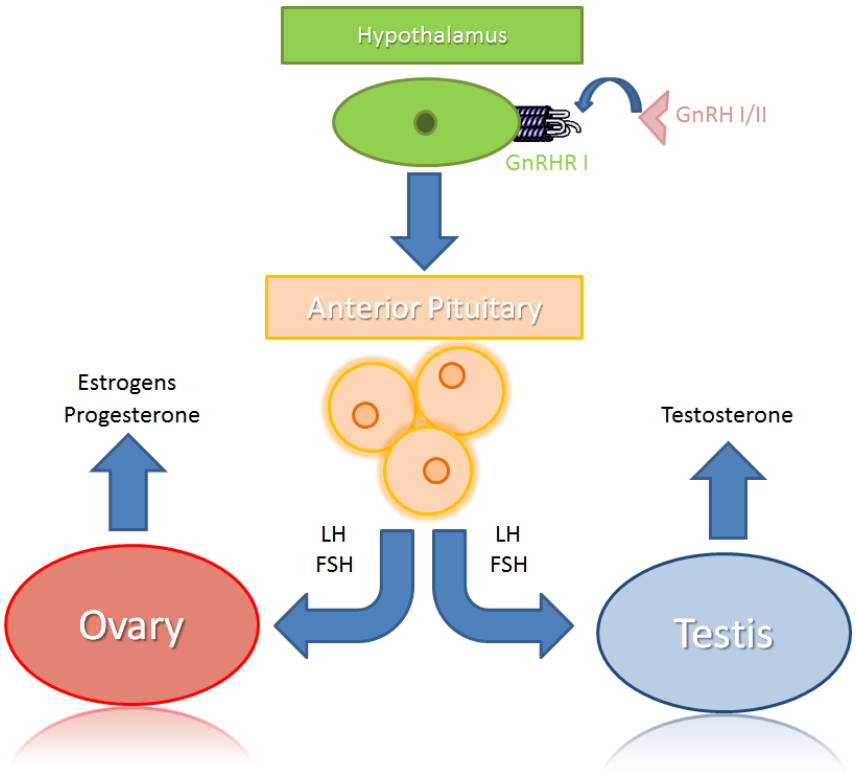


Figure 1-6. Sex steroid production is controlled by GnRH binding the cognate receptor. The GnRH neurons from the hypothalamus project into the median eminence where the GnRH is released into the portal vein allowing travel to the anterior pituitary. GnRH binds the cognate receptor and LH/FSH are released. These hormones act at the ovaries and testis to produce testosterone/estradiol.

The GnRH peptides typically bind their cognate receptor however substantial variation exists in GnRH peptide sequences (Figure 1-7). As the GnRH II cannot bind the human type two receptor it instead binds to the type one receptor. This is useful from a drug design perspective as it means more than one known active can be used to guide ligand selection.

Figure 1-7. Sequence variation in GnRH decapeptides. The sequence variation amongst the types of GnRH is mainly in residues 5-8. These determine the specificity of the hormone. Blue shaded residues are highly conserved. Image modified from reference [46].

1.5.3 Therapeutic benefits of targeting the GnRH receptor

It has been established that the GnRH receptor is critical in control of reproductive function and production of the sex steroids[51]. However it has been elucidated that GnRH-R upregulation and downregulation is found in a number of diseases. Perhaps most notable is the link between GnRH-R and hormone dependent cancers; there is evidence to suggest that treatment of prostate[37] and breast[52] cancer is possible through GnRH receptor antagonism. Other treatments are available which utilise GnRHR agonism including IVF treatment to aid conception[53] and endometriosis[54]. However to date there are still no clinically approved non-peptide drugs which target

the receptor either as agonist or antagonist. The need exists to develop a drug which is orally active and less inconvenient for patient use.

The virtual screening process involving GnRH-R is discussed in chapters 2-4.

1.6 Target selection for *in silico* screening - Post synaptic density protein 95 (PSD-95)

1.6.1 PSD-95 is a scaffolding protein which belongs to the Membrane Associated Guanylate Kinase (MAGUK) family

PSD-95 is a member of the Membrane Associated Guanylate Kinase (MAGUK) family of proteins and as such consists of three PDZ domains, an SH3 domain and a non-functional Guanylate Kinase (GK) domain[55]. Synapses are specialised structures which consist of an axon terminal and dendritic spine. These two regions are unconnected in the majority of mammals but nerve impulses must travel between them. As a result neurotransmitter is released from the axon terminal where they can then bind to the receptors on the dendritic spine and thus continue the signal by stimulating ion influx. As a result of this function the dendritic spine was found to have increased electron density in electron microscopy studies and therefore became known as the postsynaptic density (PSD)[56].

The postsynaptic density has since been found to be a complex network of 525 proteins with a combined weight of approximately 1 Million Daltons and a PSD-95 occupancy of 18%[57], where it plays important scaffolding function between a number of proteins (Figure 1-8).

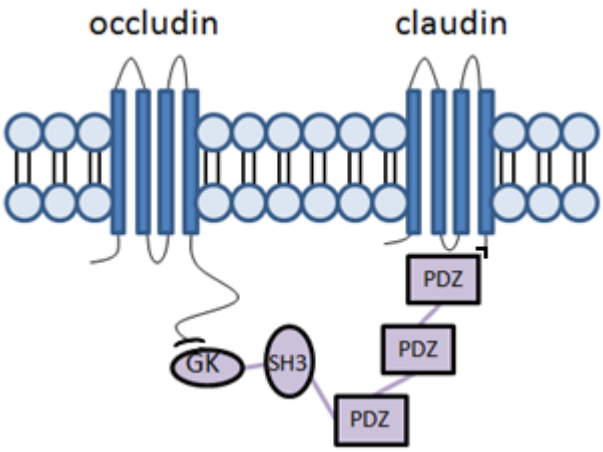


Figure 1-8. MAGUK scaffolding within the postsynaptic density. The PSD-95 protein is anchored via interactions with occludin and claudin via the GK and PDZ domains respectively. From this position the SH3 and PDZ domains are able to bind their respective motifs and act as scaffolding molecules to other proteins. Diagram adapted from reference [58].

The function of PSD-95 has recently seen greater investigation leading to the identification of a number of interesting observations. Among these was the discovery that PSD-95 was involved in the behaviour of learning and memory using PSD-95 knockout mice whereby mice deficient in the protein displayed significantly slower progress in a water maze[59]. In addition changes in expression of the PSD-95 protein is linked to a number of currently incurable neurological diseases including Parkinson's and Alzheimer's[60, 61].

1.6.2 MAGUK proteins contain novel SH3 domain structures

In the MAGUK proteins it has been found that the SH3 domain is quite unique in comparison to the canonical SH3 domain found in other proteins. Rather than the usual five beta strands this domain contains six although two of these are not part of the SH3 domain. Strand E is attached to the "HOOK" domain which is an α-helix and loop region connecting the GK and SH3 domain and strand F is part of the GK domain[62] (Figure 1-9 and 1-10).

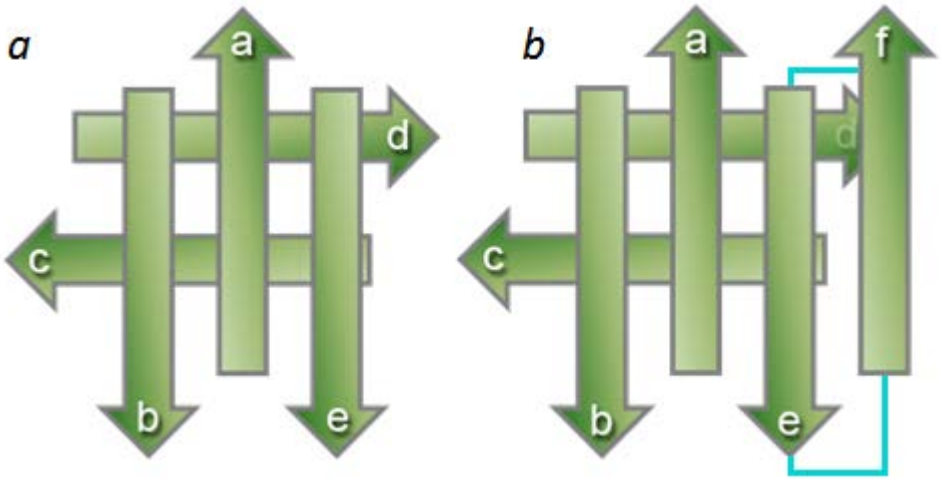


Figure 1-9. Schematic diagram comparing the beta stand orientation in the canonical SH3 domain (a) and in the MAGUK protein (b). The E and F strands in the MAGUK protein are contributed by the HOOK and GK domains respectively, the HOOK domain is shown in cyan.

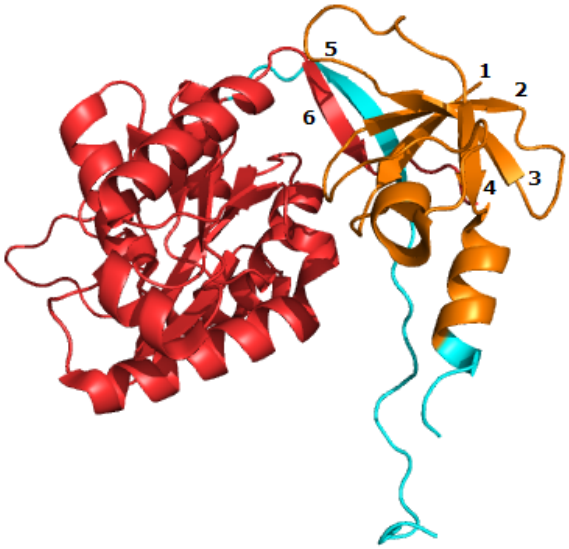


Figure 1-10. Intertwined structure of the SH3 domain in PSD-95. Spanning residues 428-709 of the protein the SH3 domain is coloured in orange and the beta strands have been numbered, the HOOK domain is coloured cyan and the GK domain in red. Note a portion of the HOOK domain is highly flexible and missing from the structure.

Further discussion on the SH3 domain structure can be found in Chapter 7.

1.6.2 PSD-95 binds to a number of therapeutically significant proteins

The N terminus of the PSD-95 protein was shown to bind directly with the SH2 domain of Src and the dopamine receptor D1 which is implicated in a number of

neurodegenerative disorders as well as the reward system of the brain[63]. It was also shown that overexpression caused an increase in the size and abundance of dendritic spines as well as enhancing the activity of glutamate receptors[64] due to interaction between PSD-95 and the glutamate receptor KA2[65] through the SH3 domain[66].

Perhaps the most interesting discovery was the association of PSD-95 with neuropathic pain in conjunction with NMDA. It was found that a shortened version of the PSD-95 protein prevented mice from exhibiting allodynia and hyperalgesia after sciatic nerve injury[67]. This makes it a very attractive target for further investigation as it may lead to the discovery of very potent new forms of pain relief.

Through the PDZ domain the protein is also able to bind to and cluster GPCR proteins to the plasma membrane. As GnRH-R has the necessary YFSL motif necessary to associate with this domain there is scope for interaction between these proteins[68].

As well as virtual screening work involving assay development and compound testing of PSD-95 hits, biochemical characterisation of the protein was also performed and is discussed in Chapter 7.

1.7 Target selection for *in silico* screening – B Lymphocyte Stimulator (BLyS)

1.7.1 BLyS is a member of the Tumour Necrosis Factor (TNF) family

BLyS (also known as BAFF, TNFSF13B or TALL1) is a cytokine involved in the maturation of B lymphocytes via interaction with the BAFF receptor (Figure 1-11) as well as TACI (transmembrane activator and calcium modulator ligand interactor) and

BCMA (B cell maturation antigen). The protein is expressed as a 285 amino acid membrane associated protein which is active in the membrane or in a cleaved soluble form after being released from the membrane via proteolytic cleavage[69]. The protein associates with TACI after it self-associates to form larger multimer complexes. Whilst it will usually form a trimer, these trimers can associate into a 60-mer virus like capsid[70] which is a requirement for signalling activity[71].

1.7.2 Maturation of B cells is an important part of immunity

A mature B cell is generated through a four stage process beginning with the differentiation of hematopoietic stem cells into pro B cells and culminating in the maturation of immature B cells to mature B cells[72]. The maturation stages occur in special germinal centres in secondary lymphoid tissues providing humoral immunity and long-lived serological memory. This process is tightly controlled to prevent production of autoantibodies whilst maintaining the required level of humoral immunity.

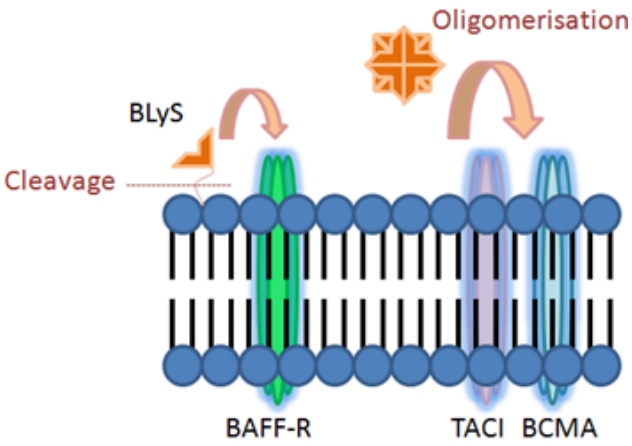


Figure 1-11. BLyS association with the cognate receptor and the TACI and BCMA receptors. Cleavage of BLyS from the membrane is possible via a furin protease, multimerisation is a requirement for activation of signalling activity with TACI and BCMA.

1.7.2 Targetting BLyS for treatment of autoimmune disease

It has been noted in the literature that elevated levels of BLyS can be found in the serum of patients suffering from a number of autoimmune diseases, including rheumatoid arthritis[73], Sjogren's syndrome[74] and systemic lupus[75]. Recently there have been clinical trials to assess the efficacy of targeting BLyS with a monoclonal antibody, Belimumab, developed by Human Genome Sciences and GSK in the treatment of systemic lupus[76]. Trials for a BLyS selective peptibody antagonist called Blisibimod (Anthera Pharmaceuticals) are also due to begin shortly, again for the treatment of lupus. Therefore there is a niche for the design of an orally active small molecule inhibitor which has the benefits of being more deliverable than the current treatments in clinical testing.

Chapter 8 contains the details of the virtual screening workflow used with BLyS as well as characterisation of the oligomeric state of the protein after extensive refolding study.

1.8 References

1. Rawlins, M.D., *Cutting the cost of drug development?* Nat Rev Drug Discov, 2004. **3**(4): p. 360-364.

2. Myers, S. and Baker, A., *Drug discovery - an operating model for a new era.* Nat Biotech, 2001. **19**(8): p. 727-730.

3. Oprea, T.I. and Matter, H., *Integrating virtual screening in lead discovery.* Current Opinion in Chemical Biology, 2004. **8**(4): p. 349-358.

4. Lindahl, E. and Sansom, M.S.P., *Membrane proteins: molecular dynamics simulations.* Current Opinion in Structural Biology, 2008. **18**(4): p. 425-431.

5. Kubinyi, H., *Chemical similarity and biological activities.* Journal of the Brazilian Chemical Society, 2002. **13**: p. 717-726.

6. Sheridan, R.P., Nilakantan, R., Rusinko, A., Bauman, N., Haraki, K.S., and Venkataraghavan, R., *3DSEARCH: a system for three-dimensional substructure searching.* Journal of Chemical Information and Computer Sciences, 1989. **29**(4): p. 255-260.

7. Humber, GL., Lee, D., Rakhit, S., and Treasurywala, MA., *Computer-assisted investigation of structure-activity relationships in serotonin-uptake blockers.* Journal of Molecular Graphics, 1985. **3**(3): p. 84-89.

8. Main, B.G., Tucker, H., and West, G.P.E.a.G.B., *3 Recent Advances in β-Adrenergic Blocking Agents*, in *Progress in Medicinal Chemistry*1985, Elsevier. p. 121-164.

9. Koes, D.R. and Camacho, C.J., *Pharmer: Efficient and Exact Pharmacophore Search.* Journal of Chemical Information and Modeling, 2011. **51**(6): p. 1307-1314.

10. Irwin, J.J. and Shoichet, B.K., *ZINC − A Free Database of Commercially Available Compounds for Virtual Screening.* Journal of Chemical Information and Modeling, 2004. **45**(1): p. 177-182.

11. Tanimoto, T.T., IBM Internal Report, 1957.

12. Duan, J., Dixon, S.L., Lowrie, J.F., and Sherman, W., *Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods.* Journal of Molecular Graphics and Modelling, 2010. **29**(2): p. 157-170.

13. Ballester, P.J., Finn, P.W., and Richards, W.G., *Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology*. Journal of Molecular Graphics and Modelling, 2009. **27**(7): p. 836-845.

14. Jenkins, J.L., Glick, M., and Davies, J.W., *A 3D Similarity Method for Scaffold Hopping from Known Drugs or Natural Ligands to New Chemotypes.* Journal of Medicinal Chemistry, 2004. **47**(25): p. 6144-6159.

15. Ballester, P.J. and Richards, W.G., *Ultrafast shape recognition for similarity search in molecular databases.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 2007. **463**(2081): p. 1307-1321.

16. Noha, S.M., Atanasov, A.G., Schuster, D., Markt, P., Fakhrudin, N., Heiss, E.H., Schrammel, O., Rollinger, J.M., Stuppner, H., Dirsch, V.M., and Wolber, G., *Discovery of a novel IKK-[beta] inhibitor by ligand-based virtual screening techniques.* Bioorganic & Medicinal Chemistry Letters, 2011. **21**(1): p. 577-583.

17. Weidlich, I.E., Dexheimer, T., Marchand, C., Antony, S., Pommier, Y., and Nicklaus, M.C., *Inhibitors of human tyrosyl-DNA phospodiesterase (hTdp1) developed by virtual screening using ligand-based pharmacophores.* Bioorganic & Medicinal Chemistry, 2010. **18**(1): p. 182-189.

18. Kellenberger, E., Rodrigo, J., Muller, P., and Rognan, D., *Comparative evaluation of eight docking tools for docking and virtual screening accuracy.* Proteins: Structure, Function, and Bioinformatics, 2004. **57**(2): p. 225-242.

19. Warren, G.L., Andrews, C.W., Capelli, A.-M., Clarke, B., LaLonde, J., Lambert, M.H., Lindvall, M., Nevins, N., Semus, S.F., Senger, S., Tedesco, G., Wall, I.D., Woolven, J.M., Peishoff, C.E., and Head, M.S., *A Critical Assessment of Docking Programs and Scoring Functions.* Journal of Medicinal Chemistry, 2005. **49**(20): p. 5912-5931.

20. Cole, J.C., Murray, C.W., Nissink, J.W.M., Taylor, R.D., and Taylor, R., *Comparing protein–ligand docking programs is difficult.* Proteins: Structure, Function, and Bioinformatics, 2005. **60**(3): p. 325-332.

21. Klon, A.E., Glick, M., Thoma, M., Acklin, P., and Davies, J.W., *Finding More Needles in the Haystack: A Simple and Efficient Method for Improving High-Throughput Docking Results.* Journal of Medicinal Chemistry, 2004. **47**(11): p. 2743-2749.

22. Jacobsson, M. and Karlén, A., *Ligand Bias of Scoring Functions in Structure-Based Virtual Screening.* Journal of Chemical Information and Modeling, 2006. **46**(3): p. 1334-1343.

23. Krier, M., Bret, G., and Rognan, D., *Assessing the Scaffold Diversity of Screening Libraries.* Journal of Chemical Information and Modeling, 2006. **46**(2): p. 512-524.

24. Lipinski, C., Lombardo, F., Dominy, B., and Feeney, P., *Experimental and computational approaches to estimate solubility and permeability in drug*

discovery and development settings. Advanced Drug Delivery Reviews, 1996. **46**(1-3): p. 3-26.

25. Ghose, A.K., Viswanadhan, V.N., and Wendoloski, J.J., *A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases.* Journal of Combinatorial Chemistry, 1998. **1**(1): p. 55-68.

26. Oprea, T.I., *Property distribution of drug-related chemical databases*.* Journal of Computer-Aided Molecular Design, 2000. **14**(3): p. 251-264.

27. Lagorce, D., Maupetit, J., Baell, J., Sperandio, O., Tufféry, P., Miteva, M.A., Galons, H., and Villoutreix, B.O., *The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections.* Bioinformatics, 2011. **27**(14): p. 2018-2020.

28. OpenEye Scientific. *http://www.eyesopen.com/filter*. [cited 2011 August].

29. Melagraki, G. and Afantitis, A., *Ligand and Structure Based Virtual Screening Strategies for Hit-Finding and Optimization of Hepatitis C Virus (HCV) Inhibitors.* Current Medicinal Chemistry. **18**(17): p. 2612-2619.

30. Bustamante, P., Muela, S., Escalera, B., and Pena, A, *Solubility Behavior and Prediction for Antihelmintics at Several Temperatures in Aqueous and Nonaqueous Mixtures.* Chemical & Pharmaceutical Bulletin, 2010. **58**(5): p. 644-649.

31. Ran, Y., He, Y., Yang, G., Johnson, J.L.H., and Yalkowsky, S.H., *Estimation of aqueous solubility of organic compounds by using the general solubility equation.* Chemosphere, 2002. **48**(5): p. 487-509.

32. Morris, J.J. and Bruneau, P.P., *Prediction of Physicochemical Properties*, in *Virtual Screening for Bioactive Molecules*2008, Wiley-VCH Verlag GmbH. p. 33-58.

33. Pharma Algorithms. *http://pharma-algorithms.com/dmso.html*. [cited 2011 August].

34. Delaney, J.S., *Predicting aqueous solubility from structure.* Drug Discovery Today, 2005. **10**(4): p. 289-295.

35. Myrdal, P.B., Manka, A.M., and Yalkowsky, S.H., *AQUAFAC 3: aqueous functional group activity coefficients; application to the estimation of aqueous solubility.* Chemosphere, 1995. **30**(9): p. 1619-1637.

36. Hapgood, J.P., Sadie, H., Van Biljon, W., and Ronacher, K., *Regulation of Expression of Mammalian Gonadotrophin-Releasing Hormone Receptor Genes.* Journal of Neuroendocrinology, 2005. **17**(10): p. 619-638.

37. Gnanapragasam, V.J., Darby, S., Khan, M.M., Lock, W.G., Robson, C.N., and Leung, H.Y., *Evidence that prostate gonadotropin-releasing hormone receptors mediate an anti-tumourigenic response to analogue therapy in hormone refractory prostate cancer.* The Journal of Pathology, 2005. **206**(2): p. 205-213.

38. Arbuckle, M.I., Komiyama, N.H., Delaney, A., Coba, M., Garry, E.M., Rosie, R., Allchorne, A.J., Forsyth, L.H., Bence, M., Carlisle, H.J., O'Dell, T.J., Mitchell, R., Fleetwood-Walker, S.M., and Grant, S.G.N., *The SH3 domain of postsynaptic density 95 mediates inflammatory pain through phosphatidylinositol-3-kinase recruitment.* EMBO Rep. **11**(6): p. 473-478.

39. Dőrner, T., *Crossroads of B cell activation in autoimmunity: rationale of targeting B cells*, 2006. p. 3-11.

40. Cancro, M.P., D'Cruz, D.P., and Khamashta, M.A., *The role of B lymphocyte stimulator (BLyS) in systemic lupus erythematosus.* The Journal of Clinical Investigation, 2009. **119**(5): p. 1066-1073.

41. Pawson, A., Morgan, K., Maudsley, S., and Millar, R., *Type II gonadotrophin-releasing hormone (GnRH-II) in reproductive biology.* Reproduction, 2003. **126**(3): p. 271-278.

42. Pawson, A.J., Maudsley, S., Morgan, K., Davidson, L., Naor, Z., and Millar, R.P., *Inhibition of Human Type I Gonadotropin-Releasing Hormone Receptor (GnRHR) Function by Expression of a Human Type II GnRHR Gene Fragment.* Endocrinology, 2005. **146**(6): p. 2639-2649.

43. Pawson, A.J. and McNeilly, A.S., *The pituitary effects of GnRH.* Animal Reproduction Science, 2005. **88**(1-2): p. 75-94.

44. Millar, R.P., *GnRH II and type II GnRH receptors.* Trends in Endocrinology and Metabolism, 2003. **14**(1): p. 35-43.

45. Davidson, J.S., *Absence of rapid desensitization of the mouse gonadotropin-releasing hormone receptor.* Biochem J, 1994. **300**: p. 299-302.

46. Millar, R.P., *GnRHs and GnRH receptors.* Animal Reproduction Science, 2005. **88**(1): p. 5-28.

47. Gault, P.M., Morgan, K., Pawson, A.J., Millar, R.P., and Lincoln, G.A., *Sheep Exhibit Novel Variations in the Organization of the Mammalian Type II Gonadotropin-Releasing Hormone Receptor Gene*, Endocrinology, 2004. p. 2362-2374.

48. Millar, R.P., Lu, Z.-L., Pawson, A.J., Flanagan, C.A., Morgan, K., and Maudsley, S.R., *Gonadotropin-Releasing Hormone Receptors.* Endocrine Reviews, 2004. **25**(2): p. 235-275.

49. Stojilkovic, S., Reinhart, J., and Catt, K., *Gonadotropin-Releasing Hormone Receptors: Structure and Signal Transduction Pathways.* Endocrine Reviews, 1994. **15**(4): p. 462-499.

50. Limonta, P., Moretti, R.M., Marelli, M.M., and Motta, M., *The biology of gonadotropin hormone-releasing hormone: role in the control of tumor growth and progression in humans.* Frontiers in Neuroendocrinology, 2003. **24**(4): p. 279-295.

51. Maruska, K.P. and Fernald, R.D., *Reproductive status regulates expression of sex steroid and GnRH receptors in the olfactory bulb.* Behavioural Brain Research. **213**(2): p. 208-217.

52. Manni, A., Santen, R., Harvey, H., Lipton, A., and Max, D., *Treatment of Breast Cancer with Gonadotropin-Releasing Hormone.* Endocrine Reviews, 1986. **7**(1): p. 89-94.

53. van Loenen, A.C.D., Huirne, J.A.F., Schats, R., Hompes, P.G.A., and Lambalk, C.B., *GnRH Agonists, Antagonists, and Assisted Conception.* Semin Reprod Med, 2002. **20**(04): p. 349,364.

54. Altintas, D., Kokcu, A., Tosun, M., Cetinkaya, M.B., and Kandemir, B., *Comparison of the effects of cetrorelix, a GnRH antagonist, and leuprolide, a GnRH agonist, on experimental endometriosis.* Journal of Obstetrics and Gynaecology Research, 2008. **34**(6): p. 1014-1019.

55. Gardoni, F. and Di Luca, M., *New targets for pharmacological intervention in the glutamatergic synapse.* European Journal of Pharmacology, 2006. **545**(1): p. 2-10.

56. Boeckers, T., *The postsynaptic density.* Cell and Tissue Research, 2006. **326(2)**: p. 409-422.

57. Chen, X., Vinade, L., Leapman, R.D., Petersen, J.D., Nakagawa, T., Phillips, T.M., Sheng, M., and Reese, T.S., *Mass of the postsynaptic density and enumeration of three key molecules.* PNAS, 2005. **102**(32): p. 11551-11556.

58. Gonzalez-Mariscal, L., Betanzos, A., and Ã Ávila, A., *MAGUK proteins: structure and role in the tight junction.* Seminars in Cell & Developmental Biology, 2000. **11**(4): p. 315-324.

59. Migaud, M., Charlesworth, P., Dempster, M., Webster, L.C., Watabe, A.M., Makhinson, M., He, Y., Ramsay, M.F., Morris, R.G.M., Morrison, J.H., O'Dell, T.J., and Grant, S.G.N., *Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein.* Nature, 1998. **396**(6710): p. 433-439.

60. Nash, J.E., Johnston, T.H., Collingridge, G.L., Garner, C.C., and Brotchie, J.M., *Subcellular redistribution of the synapse-associated proteins PSD-95*

and SAP97 in animal models of Parkinson's disease and L-DOPA-induced dyskinesia. The FASEB Journal, 2005.

61. Gylys, K.H., Fein, J.A., Yang, F., Wiley, D.J., Miller, C.A., and Cole, G.M., *Synaptic Changes in Alzheimer's Disease: Increased Amyloid-β and Gliosis in Surviving Terminals Is Accompanied by Decreased PSD-95 Fluorescence.* The American Journal of Pathology, 2004. **165**(5): p. 1809-1817.

62. McGee, A.W., Dakoji, S.R., Olsen, O., Bredt, D.S., Lim, W.A., and Prehoda, K.E., *Structure of the SH3-Guanylate Kinase Module from PSD-95 Suggests a Mechanism for Regulated Assembly of MAGUK Scaffolding Proteins.* Molecular Cell, 2001. **8**(6): p. 1291-1301.

63. Zhang, J., Vinuela, A., Neely, M.H., Hallett, P.J., Grant, S.G.N., Miller, G.M., Isacson, O., Caron, M.G., and Yao, W.-D., *Inhibition of the Dopamine D1 Receptor Signaling by PSD-95.* J Biol Chem, 2007. **282**(21): p. 15778-15789.

64. El-Husseini, A., Schnell, E., Chetkovich, D.M., Nicoll, R.A., and Bredt, D.S., *PSD-95 Involvement in Maturation of Excitatory Synapses,* Science, 2000. p. 1364-1368.

65. Mehta, S., Wu, H., Garner, C.C., and Marshall, J., *Molecular Mechanisms Regulating the Differential Association of Kainate Receptor Subunits with SAP90/PSD-95 and SAP97.* J Biol Chem, 2001. **276**(19): p. 16092-16099.

66. Garcia, E.P., Mehta, S., Blair, L.A.C., Wells, D.G., Shang, J., Fukushima, T., Fallon, J.R., Garner, C.C., and Marshall, J., *SAP90 Binds and Clusters Kainate Receptors Causing Incomplete Desensitization.* Neuron, 1998. **21**(4): p. 727-739.

67. Garry, E.M., Moss, A., Delaney, A., O'Neill, F., Blakemore, J., Bowen, J., Husi, H., Mitchell, R., Grant, S.G.N., and Fleetwood-Walker, S.M., *Neuropathic Sensitization of Behavioral Reflexes and Spinal NMDA Receptor/CaM Kinase II Interactions Are Disrupted in PSD-95 Mutant Mice.* Current Biology, 2003. **13**(4): p. 321-328.

68. Dong, C., Filipeanu, C.M., Duvernay, M.T., and Wu, G., *Regulation of G protein-coupled receptor export trafficking.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2007. **1768**(4): p. 853-870.

69. Daridon, C., Youinou, P., and Pers, J.-O., *BAFF, APRIL, TWE-PRIL: Who's who?* Autoimmunity Reviews, 2008. **7**(4): p. 267-271.

70. Liu, Y., Xu, L., Opalka, N., Kappler, J., Shu, H.-B., and Zhang, G., *Crystal Structure of sTALL-1 Reveals a Virus-like Assembly of TNF Family Ligands.* Cell, 2002. **108**(3): p. 383-394.

71. Liu, Z. and Davidson, A., *BAFF inhibition: A new class of drugs for the treatment of autoimmunity*. Experimental Cell Research. **317**(9): p. 1270-1277.

72. Tangye, S.G., Bryant, V.L., Cuss, A.K., and Good, K.L., *BAFF, APRIL and human B cell disorders.* Seminars in Immunology, 2006. **18**(5): p. 305-317.

73. Cheema, G.S., Roschke, V., Hilbert, D.M., and Stohl, W., *Elevated serum B lymphocyte stimulator levels in patients with systemic immune–based rheumatic diseases.* Arthritis & Rheumatism, 2001. **44**(6): p. 1313-1319.

74. Varin, M.-M., Le Pottier, L., Youinou, P., Saulep, D., Mackay, F., and Pers, J.-O., *B-cell tolerance breakdown in Sjögren's Syndrome: Focus on BAFF*. Autoimmunity Reviews. **9**(9): p. 604-608.

75. Fawzy, S.M., Gheita, T.A., El-Nabarawy, E., El-Demellawy, H.H., and Shaker, O.G., *Serum BAFF level and its correlations with various disease parameters in patients with systemic sclerosis and systemic lupus erythematosus.* The Egyptian Rheumatologist. **33**(1): p. 45-51.

76. Navarra, S.V., Guzmán, R.M., Gallacher, A.E., Hall, S., Levy, R.A., Jimenez, R.E., Li, E.K.M., Thomas, M., Kim, H.-Y., León, M.G., Tanasescu, C., Nasonov, E., Lan, J.-L., Pineda, L., Zhong, Z.J., Freimuth, W., and Petri, M.A., *Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial.* The Lancet, 2011. **377**(9767): p. 721-731.

Chapter 2. Computational and biochemical protocols

2.1 Computational tools and approaches

The number of protein crystal structures covers approximately 10% of the known protein sequences[1, 2]. This discrepancy causes a continued need for the ability to generate protein structure using computational tools. Typically protein modelling is performed using known structures with high sequence identity to the protein of interest; however other structural information such as NMR data, image picking from electron microscopy or just using rudimentary rules of secondary structure organisation can theoretically be used to generate a model structure.

In this chapter the computational tools that were used to generate homology models for the GnRH receptor using sequence homology to the rhodopsin and β_2 adrenergic receptor are described. Docking programs and algorithms used to identify potential small molecule ligands are presented with the biochemical assays that were performed to test the inhibitors.

2.2 MODELLER

MODELLER[3] is a software suite designed to produce high quality three dimensional homology models using a template structure. Spatial constraints between Cα atoms, hydrogen bonds and dihedral angles which are calculated from the template structure are applied to the sequence to be modelled (Figure 2-1). The steps used in the MODELLER process are described below.

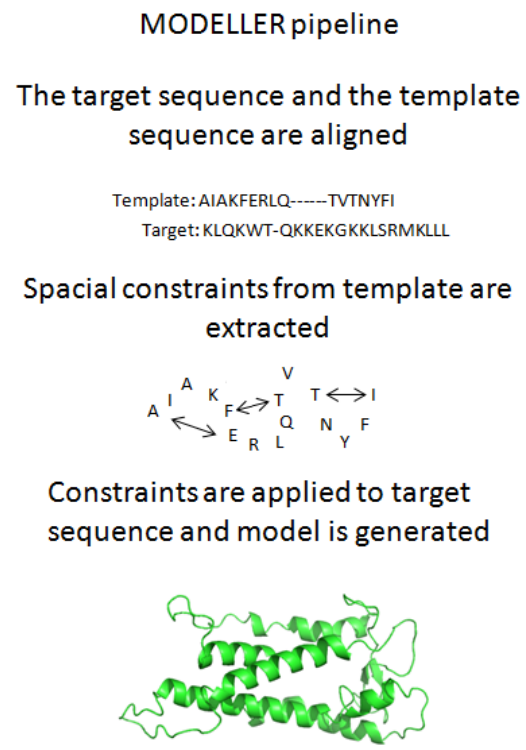


Figure 2-1. An overview of the model structure building process in the MODELLER program.

2.2.1 Alignment of target and template sequences in MODELLER

Sequence alignment in MODELLER can be executed by a Needleman-Wunsch[4] method which involves a global alignment of two sequences in an array where a numerical score is given to the similarity between two residues. The final alignment produced is the pathway through the array with the highest score. An alternative approach uses the Smith-Waterman algorithm[5] which is a local alignment method rather than a global one. This is slightly different from the Needleman-Wunsch algorithm in that when the array is created and the residue similarities are scored the score is based on sequences of any length, therefore gap penalties are introduced to prevent excessive insertions. Assuming two sequences are highly similar then both algorithms will give similar results, however when two sequences have low similarity the Smith-Waterman algorithm is the preferred choice[6].

Initially alignments were produced using the Smith-Waterman algorithm in MODELLER, however these alignments did not correctly align conserved motifs between the G protein coupled receptor sequences. Alignments were instead generated using a pairwise alignment with the BLOSUM matrix in ClustalW[7] which is more robust when aligning sequences of low similarity. This is achieved by assigning higher scores to rarer amino acids that are found paired over more common ones. For example a Leu/Leu pair scores +4 whilst a Trp/Trp pair scores +11 because statistically Trp is a less common amino acid[8]. It is notable that matrices such as PHAT[9] have been designed specifically for membrane bound proteins, however it has been shown that ClustalW is equally applicable in membrane bound protein sequence alignment[10].

2.2.2 Calculating spacial restraints in the template structure

After aligning the two sequences (or after two pre-aligned sequences are provided as in the above case) MODELLER calculates spatial restraints from the coordinate file (in the PDB file format) to apply to the model.

The calculation of spatial constraints involves measuring a number of distances and bond angles. These include template C_α- C_α distances, backbone dihedrals (φ/ψ), sidechain dihedrals and van der Waals contacts. All of these constraints are automatically calculated by the software and stored for the next step.

2.2.3 Generating the target model from spatial restraints of a homologous structure.

Distance constraint rules used in MODELLER are derived using a structural database containing alignments of homologous proteins[11]. Model structures are generated by

using a Gaussian function (Equation 1) to calculate a probability density function for each of these spacial constraints based on the template structure and the similarity to the target structure. As an example the calculation of the probability density function for C_α- C_α distance is shown in Equation 1.

$$p(d'/d) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{d'-d}{\sigma}\right)^2\right]$$ Equation 1

Where *d'* and *d* are the two equivalent C_α- C_α distances in the template and unknown structures respectively, σ is the standard deviation of the distribution of both distances. This is calculated from a polynomial which has three main dependencies; the similarity between the two proteins, the solvent exposure of the two residues spanning the distance and the proximity to the nearest gaps in sequence alignment between both proteins. A typical value is 1Å. Thereafter the spatial constraints and the CHARMM22 energy terms are combined into an objective function which is optimised in Cartesian space. The optimisation is performed using molecular dynamics with simulated annealing.

MODELLER outputs a selection of structures which are scored by Discrete Optimized Protein Energy (DOPE) to suggest the best model which completes the modelling pipeline in MODELLER.

The DOPE is an atomic distance-dependent statistical potential calculated from a sample of native protein structures. This uses probability theory to quantify if the differences between the template structure and the model are energetically favourable given the probability functions which were used to build the model.

Further refinement of the model is possible using other MODELLER functions including additional loop refinement which was required in the creation of GnRH-R structures. Because the sequence homology is lowest in loop regions between GPCR proteins the suggested loop regions are often incorrect. Sometimes this can be resolved by comparing the sequence to a protein database which may contain the required structural information for the required sequence which can then be optimised using an energy function. This type of approach is normally only useful for loop regions shorter than 6 residues, which is approximately half the size of a GPCR loop. Alternatively the loop can be refined *ab initio* using the loop.py module of MODELLER which runs the loop sequence through a number of relaxation steps; these allow for close atom contact before a molecular dynamics simulation which features rapid heating to 1000K and then a gentle cooling stage back to just above room temperature followed by a final relaxation gradient[12].

2.3 ProsaWeb

A number of algorithms exist which are designed to generate 3D protein structure by homology modelling. Each modelling program typically has an inbuilt scoring system which assesses model quality by the minimum energy of the protein in the model conformation. Whilst this gives a general indication of the model quality it does not provide any insight into the quality of folding that is present in the model. In order to achieve this an independent method of testing was sought and in this case ProsaWeb[13, 14] was selected for this purpose.

The ProsaWeb software was originally intended as a means of validating the accuracy of X-ray crystal structures; however, the method is also useful for

assessment of model structures. ProsaWeb runs on a webserver which is freely accessible and accepts model structures for error analysis in the standard PDB format. The program does not calculate model accuracy based on violation of steric principles such as atom close contact, its only function is to identify misfolded regions. These were developed using a "hide and seek" approach of picking out correctly folded regions from 50000 decoys[14]. Only the C_α atoms are used in the evaluation procedure; the energy of the structure as well as solvent exposure of each residue is calculated and used to derive a z score and residue energy. The z-score is a measure of the deviation of the total energy from the structure when compared to an energy distribution derived from the sampling random conformations. The z score alone is only meaningful when compared against the z score of proteins with similar size. The residue energies are calculated in linear fragments of 40 amino acids rather than as individual residues. This reduces large individual fluctuations in residue energy which the authors claim is not useful when investigating protein folds.

2.4 LIDAEUS

LIDAEUS (Ligand discovery at Edinburgh University)[15, 16] is a high throughput rigid body docking program which was written as a series of connected pipelines between 4 modular docking and scoring components (Figure 2-2).

Figure 2-2. LIDAEUS docking pipeline. Preen adds atom type information and assigns partial charges. The Pose module is where the docking of the compound into the protein occurs, it will attempt to fit the molecule into the site points and outputs instances where atoms from site points and compounds match. Score assesses the potential energies of interactions between matched atoms and generates an overall score which is then sorted from best to worst by Sort.

Before a protein can be submitted to LIDAEUS there must first be a set of energy maps defined for the target protein which are generated by using probe atoms around the space occupied by a ligand molecule docked into the target area. Three energy maps in total are required which contain information on hydrogen bonding, hydrophobic interactions and van der Waals attractions. The van der Waals attractions are defined using a Lennard Jones potential which describes the interaction between two uncharged molecules or atoms at a given distance.

The map containing hydrogen bond donors and acceptors use complimentary probes to perform the respective calculations. These are weighted to favour specific hydrogen bond angles depending on the nature of atoms that are involved.

LIDAEUS is able to use these three maps to approximate the energy of interaction of any ligand molecule that is docked within the ligand binding site specified earlier. Obviously large proteins may contain more than one binding site. Docking millions of small molecule fragments into a protein site which is not the target is wasteful in CPU time; therefore the maps also serve to direct LIDAEUS docking activity to defined regions of the target protein. The site points used in the rhodopsin based model of GnRH-R are shown in Figure 2-3.

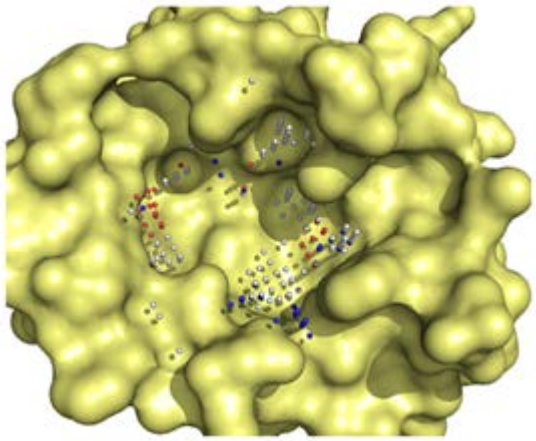


Figure 2-3. Site-points created for the rhodopsin based GnRH receptor model. White points represent hydrophobic contacts, red points indicated hydrogen bond acceptors and blue points are hydrogen bond donors.

2.4.1 The preen stage

Preen is the first stage in the LIDAEUS pipeline and is the module which reads in the small molecule from the users compound library formatted in an SDF file. During the molecule reading process preen adds additional atom type information, primarily this includes whether the atom is in a ring, the partial charges on the atom and also any missing hydrogens are added.

2.4.2 The pose stage

In this step the small molecule which has been atom typed by preen is fitted into the target binding site of the protein as directed by the site points file. The program cycles through as many orientations as possible but without manipulating the rotational bonds within the small molecule. Because LIDAEUS is a rigid body docking program the ligand can only be manipulated by rotation and translation movements which are constrained by the site points provided. A successful pose is established by an atom position being docked within a certain tolerated distance of the site points; by default the first four atoms must be within 0.02 Å, 0.04 Å, 0.06 Å

and 0.08 Å. If the ligand matches these requirements then the ligand coordinates are saved into the ligand file and this is piped into the Score module.

2.4.3 The score stage

Ligands entering the scoring stage have their associated atom type values and the coordinates which were found to fit the defined fitting tolerance to the site points. The score module then takes the successful pose and uses the energy maps to calculate the energy contribution from each atom, giving a value in kcal /mol. The binding pose is then refined by applying energy minimisation before further rotation and translation is implemented to move the ligand into a more favourable position. During this ligand movement there is no site point involvement as the tolerance values associated with this are already considered to be met. In the most energetically favourable pose the binding energies contributed from van der Waals, hydrophobic and hydrogen bond interactions are used to make the score which will be used to rank the docked pose against the other molecules in the library.

2.4.4 The Sort stage

This is the simplest operation in the LIDAEUS workflow. Sort need only read the score information which is appended in the previous step and organise the top 1000 compounds (or the user selected number) into numerical order. As the scores are displayed as values of kcal /mol the more negative the number the better the perceived pose interaction; in LIDAEUS docking a score which is less than -20 kcal /mol is considered to be a sensible cut off point.

2.5 AutoDock vina

AutoDock vina[17] is a docking program which has the advantage of being able to treat both the ligand and protein as flexible structures. To start a docking run with AutoDock vina a little preparation is necessary; first the ligand and protein must be saved in the PDBQT file format and secondly the docking space must be defined.

The PDBQT file format is unique to AutoDock and is generated using AutoDock Tools in the MGLTools package. The ligand and protein can be read from a PDB file after which the rotatable bonds which the program can identify automatically will be highlighted (Figure 2-4); these can be manually altered to make certain bonds more or less flexible depending on user preference. After rotatable bonds on the ligand and protein are suitable, both files are saved as the prerequisite PDBQT file. The next step is to define the search space which will be used in the docking procedure in the same way that the site points in LIDAEUS direct rigid body docking but without the associated interaction type information.

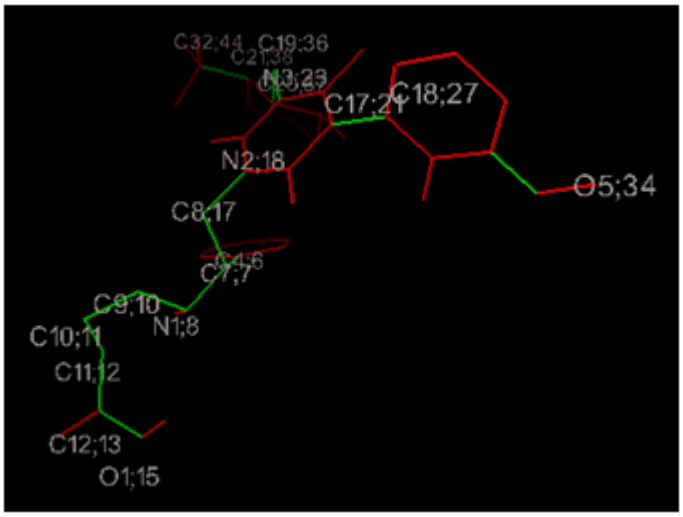


Figure 2-4. Defining rotatable bonds of a sample ligand structure using AutoDockTools. Bonds which are highlighted in green are rotatable, those coloured red are not.

The docking area sampling method is fairly simplistic, and involves manipulating a coloured box into the area that is to be included as a docking stage. The volume can be as large or as small as required (Figure 2-5); however there is an exponential increase in computation time as the volume of the box increases. Therefore a focussed search is preferable if the ligand binding site has already been biophysically established.

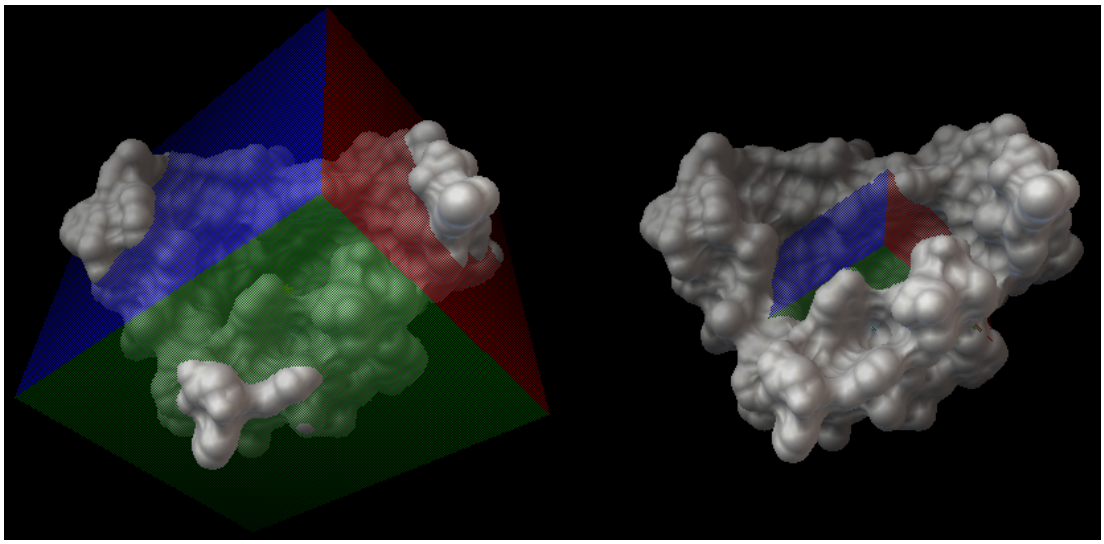


Figure 2-5. Defining docking area of the target protein using AutoDockTools. The docking area is able to encompass an area as large as the user prefers; however, a search space greater than 27000Å^3 will prompt a warning from AutoDock vina.

Once the docking area has been defined the coordinates of the grid box must be noted in order to be included in a configuration file which is required to supply to AutoDock vina via the command line. The configuration file also specifies the ligand molecule, the protein to dock into and the exhaustiveness of the search which can be set between 1 and 8, where 8 is the most exhaustive search.

When AutoDock vina is performing a docking experiment it performs a number of *runs* consisting of sequential *steps*. Each *step* involves a random perturbation of the ligand-protein conformation followed by an optimisation step (using the Broyden-

Fletcher-Goldfarb-Shanno algorithm); after optimisation the *step* is either accepted or rejected. Each optimisation consists of a number of *evaluations* with the scoring function as well as movement in the *position-orientation-torsions* coordinates. The number of *evaluations* that are performed are determined by convergence; whilst the number of *steps* in the *run* are determined heuristically depending on ligand size and flexibility. The number of *runs* are set by the exhaustiveness number that is selected in the configuration file.

Increasing the exhaustiveness has a net increase on computational time; a setting of 6 gives a reasonable trade off in computational time over increases in docking accuracy.

2.6 EDULISS

The Edinburgh University ligand selection system (EDULISS)[18] is a compound database which contains over 4 million unique compounds from the catalogues of 26 chemical companies. EDULISS is more than a concatenated searchable compound collection. Each compound is associated with 1600 pre-calculated descriptors; these descriptors offer a wealth of biophysical information beyond the simplistic molecular weight and chemical formulae data that is often the limit of other databases.

In this case EDULISS stores topological, geometrical, physicochemical and toxicological descriptors for each compound; this allows a more diverse selection of queries to be used to pre-select smaller families of molecules prior to screening. There are unique descriptors calculated for each compound which describe interatomic distances generally within the molecule as well as between hydrogen bond donors and acceptors, halogens, phosphorus and sulphurs. These allow

searching using ultrafast shape recognition which will be discussed in greater detail further in this chapter; but searching by pharmacophore is also possible using this method which is integrated into the EDULISS website (http://eduliss.bch.ed.ac.uk/test/index.jsp)

The statistics for the proportion of compounds which meet common drug likeness criteria are shown in Figure 2-6.

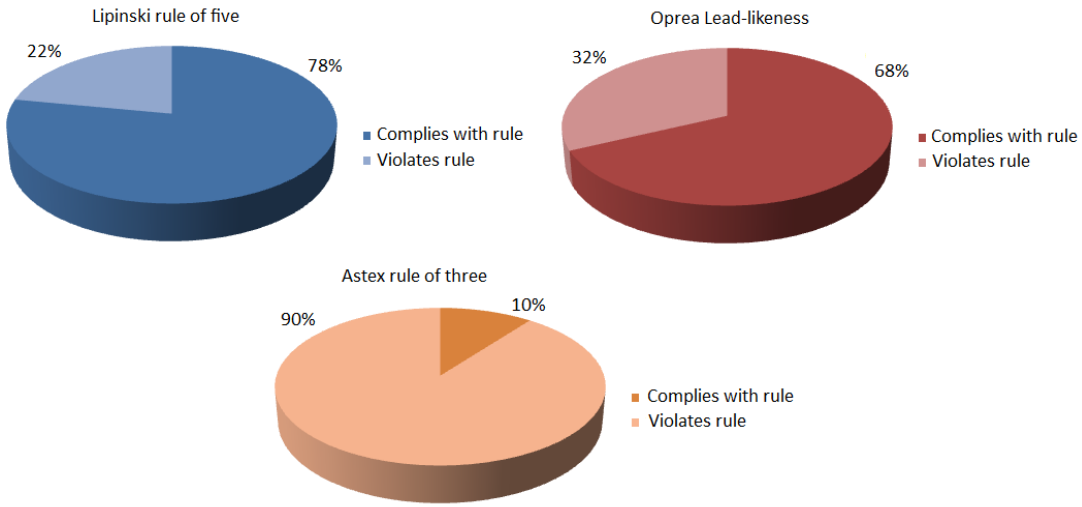


Figure 2-6. Summary of the drug likeness criteria met by the compounds in the EDULISS database. Almost 75% of the database pass both the Lipinski and Oprea criteria. Whilst the view could be taken that the whole database should pass as many drug likeness criteria as possible it should be considered that some diversity in the scaffolds are a useful method of finding novel ligands.

2.7 Compound clustering by Tanimoto coefficient

Molecular similarity is a key tool which has been used to cut the massive costs associated with developing new products within the pharmaceutical industry[19, 20]. A number of methods exist to derive similarity relationships between compounds in 1D, 2D and 3D formats.

1D similarity could be considered a “low resolution” similarity matrix as it is typically based on simple information like atom count, molecular weight, logP and chemical formulae data.

2D similarity is the most common form of similarity searching, it is relatively computationally inexpensive compared to 3D similarity searching but offers a far greater depth compared to 1D searches. A fingerprint for the 2D structure can be generated in seconds using the Daylight fingerprint software[21] although a number of other fingerprinting algorithms exist.

3D similarity has the advantage of containing the most information for similarity searching; however, it has disadvantages that 2D methods do not. The most costly implementation of 3D similarity is the requirement for the generation of flexible conformers which allow alignment against query structures. Given that there has been no real advantage of using 3D similarity over 2D[22, 23] the use of 3D descriptors has generally fallen out of favour.

After a method of similarity has been agreed the next choice is deciding which similarity scoring scheme should be used to compare structures. A number of coefficients exist which can be used to assess similarity; they are summarised in Table 2-1.

Table 2-1. Common similarity coefficients used in similarity searching. Modified from reference[24].

Name	Formula	Value range
Tanimoto	$S_{AB} = \frac{c}{a+b-c}$	0 to 1
Dice	$S_{AB} = \frac{2c}{a+b}$	0 to 1
Cosine	$S_{AB} = \frac{c}{\sqrt{ab}}$	0 to 1
Euclidean distance	$D_{AB} = \sqrt{a+b-2c}$	0 to N
Manhattan /Euclidean distance	$D_{AB} = a+b-2c$	0 to N

In Table 2-1 the number of ON bits which are in structure A are substituted into term *a*, the ON bits in structure B are substituted into *b* and the ON bits common to both A and B are substituted into term *c*. Generally if an active and an unknown compound are found to share 0.85 similarity by a Tanimoto coefficient then there is an 80% chance that the unknown compound will also have activity[25].

Despite the choice in similarity coefficients the most widely recognised is the Tanimoto coefficient. Its widespread use may be a consequence of having a "molecular size normalisation in the denominator" [26] which allows it to still be effective in work with small molecules. Other methods can show bias to larger molecules because they have more ON bit strings to match to. In a recent review of similarity methods it was found when using one coefficient alone the Tanimoto method was the most favourable, although complimentary results were obtained

when using up to 6 different coefficients in a screen[27]. When using a Tanimoto coefficient compound similarity in relation to a query molecule is ranked between 0 and 1. Scores closer to 0 indicated increasing dissimilarity between the query molecule and the compared structure; however this does not mean that scores of 1 indicate that both structures are identical, only the fingerprints are. It is also notable that the Tanimoto coefficient along with the Dice and Cosine measure the direct similarity between structures, the Euclidean and Manhattan/Euclidean coefficients measure the distance. The Tanitmoto coefficient value can however be converted to a distance measurement using the Soergel distance which simply subtracts the tanimoto value from 1 and is used in the clustering method described in this work.

After the top 1000 compounds were output from LIDAEUS a Lipinski ruleset was applied to further filter the set. The top 100 compounds were then clustered by similarity according to a Tanimoto coefficient before being converted to distance using the Soergel formula and plotted on a distance tree using an R script. A simple clustering example with the Soergel formula using a small dataset of 10 compounds is shown in Figure 2-7.

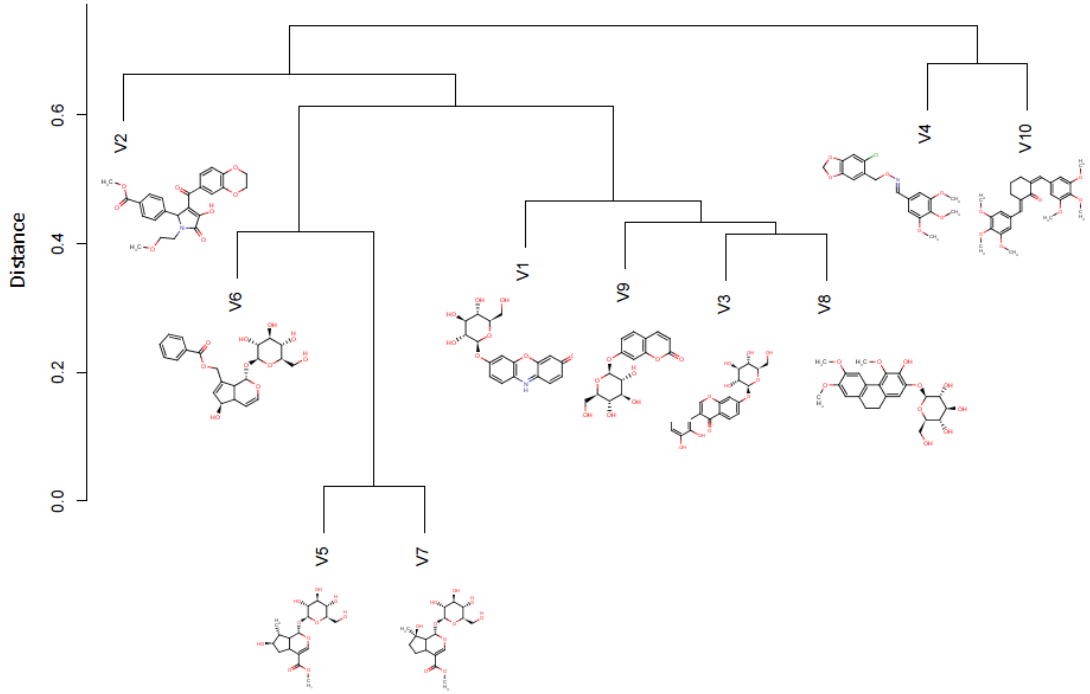


Figure 2-7. A clustering of 10 compounds by Tanimoto coefficient.

2.8 UFSRAT

Ultrafast shape recognition (UFSR)[28] was briefly described in chapter 1, in this section the basic principles will be expanded upon with a worked example on how ultrafast shape recognition with atom types (UFSRAT) can be implemented in the search for GnRH-R inhibitors.

Using UFSR as a similarity scoring system returns a value between 0 and 1; however unlike the Tanimoto coefficient, when a score of 1 is given it indicates a perfect match. The process follows two steps which simply calculate the molecular descriptors for the query structure and the unknown and then compares them.

The descriptors used in UFSR can be grouped into four sets of three, so a single molecule has twelve descriptors in total. Calculation of the descriptors follows a

logical pattern that starts with selecting point A which is the centroid of the molecule.

Next a list of Euclidean distances of all atoms to A is generated. This distribution of distances has mean, variance and skew values to A which are used to create the first three of the twelve descriptors. The atom which is closest to the centroid A is selected, let this be point B. The Euclidean distances of all atoms to B are compiled and the mean variance and skew are calculated; these account for descriptors 4 to 6.

The furthest atom from B is then selected, let this be point C and again all Euclidean distances to this atom are calculated and used to generate the mean, variance and skew data that are used in descriptors 7 to 9.

Finally, the atom which is furthest from atom C is selected, the Euclidean distances are calculated and the mean variance and skew data are used in descriptors 10 to 12.

The scoring system is obtained using the equation shown below.

$$S_{qi} = \left(1 + \frac{1}{12}\sum_{l=1}^{12}\left|M_l^{\,q} - M_l^{\,i}\right|\right)^{-1}$$

In this example the score (S) from comparing the query (q) molecule against the ith structure (i) is given by substituting the descriptor numbers for the query and ith compound into the $M_1^{\,q}$ and $M_1^{\,i}$ terms respectively.

This technique is very powerful and delivered results which were not only closer in similarity to the query structure than EShape3D but was also 1546 times faster at returning the results[28].

This method was further developed by Steven Shave, a colleague in the group, by adding information on atom types into the descriptor terms which are expanded from 12 terms to 48. This has the advantage of not only including information on all atoms but also being able to pick out hydrophobic atoms and hydrogen bond donor and acceptors.

As an example a similarity search using the structure of the non-peptide antagonist Elagolix was performed. This compares EDULISS compounds from the Chembridge, Maybridge, Pubchem, Sigma and Specs catalogues against that of Elagolix. The results are shown in Figure 2-8.

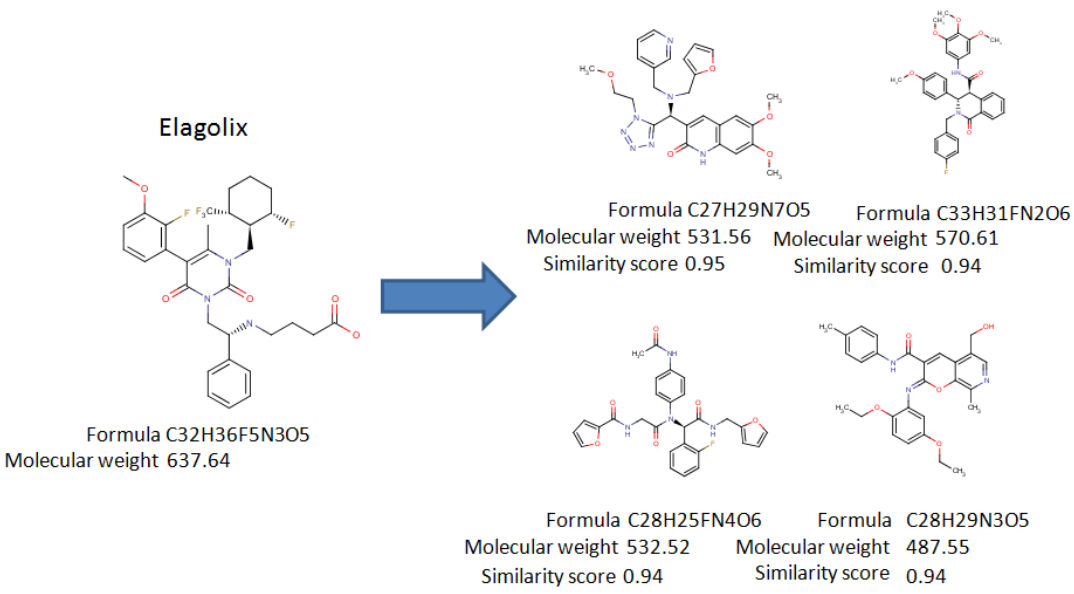


Figure 2-8. Similarity searching the EDULISS database using UFSRAT and the Elagolix structure.

2.9 Surface Triplet Propensities (STP)

Identifying ligand binding sites on proteins is a problem which has been addressed by a number of alternative methods. The most common approach relies on

identifying ligand binding sites by searching for surface cavities[29, 30]; this is fairly successful however difficulties occur when there are more than one cavity on the protein of similar depth. An alternative technique which uses the amino acid profile of a number of protein binding sites was developed using 804 ligand binding interactions[31]. This was used to create an algorithm which is not concerned with pocket depth but instead the residues which create the pocket.

STP is a protein binding site prediction program developed by Wissam Mehio[32] which operates in a similar way. The surface of the protein is analysed with a probe that is the size of a water molecule and any group of 3 atoms that can be simultaneously touched by the probe are outputted as a triplet along with the atomic group information. 14 atomic groups exist to describe the properties of that group.

An example of STP scoring the binding pocket of the rhodopsin derived model of GnRH-R is shown in Figure 2-9.

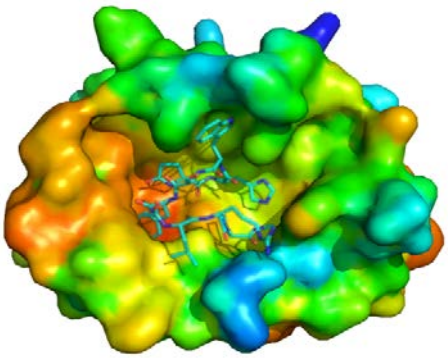


Figure 2-9. Scoring of the surface of the binding pocket of GnRH-R with STP. STP was able to correctly score the pocket where the GnRH I hairpin associates. The program also scored a number of locations on the outer surface highly, this may be a result of the protein being membrane bound.

2.10 Biochemical assays techniques

2.10.1 Culture of COS-7 cells

COS-7 cells were cultured in 162cm^2 cell culture flasks at 37 °C, 5% CO2 in complete medium (DMEM supplemented with 10% fetal calf serum). The cells were passaged once per week with cells being split approximately 1 in 20.

2.10.2 Transfection of COS-7 cells with P41 plasmid

10 µg of plasmid DNA was added to each 0.4 cm electroporation cuvette and kept on ice. Each 162cm^2 flask of COS-7 cells was split between 2 cuvettes in the following way; each flask of cells was washed with 20mls of PBS (37^{o}C, Ca^{2+}, Mg^{2+} free) twice then the cells were detached by adding 2 ml of trypsin/EDTA and incubating for 5 minutes at 37 ^{o}C.

8mls of complete medium was added to each flask and the cells were resuspended in the medium by pipetting. Resuspended cells were transferred into 50-ml sterile polypropylene centrifuge tubes and centrifuged at 2000 rpm for 3mins. The medium was aspirated and the cells in each 50 ml tube were then resuspended in 25 ml of prechilled electroporation DMEM (serum, glutamine and antibiotic free DMEM).

The cells were pelleted again and the electroporation DMEM was aspirated. Cells were then resuspended in 3.5 ml of prechilled electroporation DMEM. 0.7 ml resuspended cells were added to each electroporation cuvette and mixed with the plasmid DNA by pipetting. Cells were pulsed at 960µF (capacitance) and 0.22kV in a Bio-Rad Gene Pulser MXCell then left at room temperature for 10 minutes before being mixed with 26 mls of complete media and seeded into 12 well plates.

2.10.3 Whole cell binding assay

Transiently transfected COS-7 cells were incubated at 37 ^{o}C in complete medium for 48 hours prior to assay. Cells were checked for confluence and then incubated with fresh complete medium for 30 minutes before any compounds were added. All compounds were diluted to their final concentration into ice cold DMEM supplemented with 0.1% BSA and radiolabelled ^{125}I-[His^{5}, D-Tyr^{6}] –GnRH (100000 counts /ml) which has a very high affinity for the GnRH receptor[33]. Antagonist and peptide mixtures were incubated with the cells for 4 hours at 4 ^{o}C.

Cells were finally washed twice with ice cold PBS and solublised by shaking with 0.1M NaOH for 20 minutes at room temperature. Solubilised cells were then pipetted into 12 mm tubes and counted in a multigamma counter (Perkin Elmer Wallac).

2.10.4 Inositol phosphate (IP3) assay

24 hours after transfection the medium is replaced with 1ml inositol-free special DMEM (Gibco) containing 1% FCS, glutamine, antibiotics, and 1µl/ml ^{3}H-myo-inositol. Prior to incubation with compounds the special DMEM is aspirated and the cells are washed by incubating with HEPES-DMEM/0.1% BSA, 10mM LiCl for 30 minutes at 37 ^{o}C. This is then aspirated and 500 µl HEPES-DMEM/0.1% BSA, 10mM LiCl and 5% DMSO, with the appropriate final concentration of antagonist compound, is added to the cells and incubated for 30 minutes at 37 ^{o}C. 3nM of GnRH I is then added to each well, with the exception of the basal controls, and incubated for another 30 minutes.

Cells are lysed using 10mM formic acid added to each well and incubated at 4 ^{o}C for 30 minutes. Cell lysate is then added to AG 1-X8 resin in 12 mm plastic tubes and

briefly vortexed; once the resin settles the supernatant is aspirated and water is added, vortexed and aspirated. 60mM ammonium formate/5mM sodium tetraborate is then added, vortexed and then aspirated; thereafter 1M ammonium formate/0.1M formic acid is added then vortexed. The supernatant was then pipetted into scintillation vials where scintilint was added (Optiphase HiSafe 3) and counted using a beta counter.

2.11 References

1. Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., and Zardecki, C., *The Protein Data Bank*, in *Acta Crystallographica Section D*2002. p. 899-907.

2. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M., *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003*, 2003. p. 365-370.

3. Martí-Renom, M.A., Stuart, A.C., Fiser, A., Sánchez, R., Melo, F., and Šali, A., *Comparative Protein Structure Modeling Of Genes And Genomes*, 2000. p. 291-325.

4. Needleman, S.B. and Wunsch, C.D., *A general method applicable to the search for similarities in the amino acid sequence of two proteins.* Journal of Molecular Biology, 1970. **48**(3): p. 443-453.

5. Smith, T.F. and Waterman, M.S., *Identification of common molecular subsequences.* Journal of Molecular Biology, 1981. **147**(1): p. 195-197.

6. Pierri, C.L., Parisi, G., and Porcelli, V., *Computational approaches for protein function prediction: A combined strategy from multiple sequence alignment to molecular docking-based virtual screening.* Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics. **1804**(9): p. 1695-1712.

7. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G., *Clustal W and Clustal X version 2.0*, 2007. p. 2947-2948.

8. Eddy, S.R., *Where did the BLOSUM62 alignment score matrix come from?* Nat Biotech, 2004. **22**(8): p. 1035-1036.

9. Ng, P.C., Henikoff, J.G., and Henikoff, S., *PHAT: a transmembrane-specific substitution matrix.* Bioinformatics, 2000. **16**(9): p. 760-766.

10. Forrest, L.R., Tang, C.L., and Honig, B., *On the Accuracy of Homology Modeling and Sequence Alignment Methods Applied to Membrane Proteins.* Biophysical Journal, 2006. **91**(2): p. 508-517.

11. Sali, A., Matsumoto, R., McNeil, H.P., Karplus, M., and Stevens, R.L., *Three-dimensional models of four mouse mast cell chymases. Identification of proteoglycan binding regions and protease-specific antigenic epitopes*, 1993. p. 9023-9034.

12. Fiser, A., Do, R.K.G., and Šali, A., *Modeling of loops in protein structures.* Protein Science, 2000. **9**(9): p. 1753-1773.

13. Wiederstein, M. and Sippl, M.J., *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.* Nucleic Acids Research, 2007. **35**(suppl 2): p. W407-W410.

14. Sippl, M., *Recognition of errors in three-dimensional structures of proteins.* Protein Science, 1993. **17**: p. 355-362.

15. Shave, S.R., Taylor, P., Walkinshaw, M., Smith, L., Hardy, J., and Trew, A., *Ligand discovery on massively parallel systems.* IBM Journal of Research and Development, 2008(52): p. 57-67.

16. Taylor, P., Blackburn, E., Sheng, Y., Harding, S., Hsin, K.-Y., Kan, D., Shave, S., and Walkinshaw, M., *Ligand discovery and virtual screening using the program LIDAEUS.* British Journal of Pharmacology, 2008(153): p. S55–S67.

17. Trott, O. and Olson, A.J., *AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* Journal of Computational Chemistry, 2010. **31**(2): p. 455-461.

18. Hsin, K.-Y., Morgan, H.P., Shave, S.R., Hinton, A.C., Taylor, P., and Walkinshaw, M.D., *EDULISS: a small-molecule database with data-mining and pharmacophore searching capabilities.* Nucleic Acids Research, 2011 **39**(1).

19. DiMasi, J.A., Hansen, R.W., and Grabowski, H.G., *The price of innovation: new estimates of drug development costs.* Journal of Health Economics, 2003. **22**(2): p. 151-185.

20. Bender, A. and Glen, R.C., *Molecular similarity: a key technique in molecular informatics.* Organic & Biomolecular Chemistry, 2004. **2**(22): p. 3204-3218.

21. Daylight Chemical Information Systems. *http://www.daylight.com/*. [cited 2010 November].

22. Matter, H. and Pötter, T., *Comparing 3D Pharmacophore Triplets and 2D Fingerprints for Selecting Diverse Compound Subsets.* Journal of Chemical Information and Computer Sciences, 1999. **39**(6): p. 1211-1225.

23. Sheridan, R.P. and Kearsley, S.K., *Why do we need so many chemical similarity search methods?* Drug Discovery Today, 2002. **7**(17): p. 903-911.

24. Maggioni, M., Santambrogio, M.D., and Liang, J., *GPU-accelerated Chemical Similarity Assessment for Large Scale Databases.* Procedia Computer Science. **4**(0): p. 2007-2016.

25. Martin, Y.C., Kofron, J.L., and Traphagen, L.M., *Do Structurally Similar Molecules Have Similar Biological Activity?* Journal of Medicinal Chemistry, 2002. **45**(19): p. 4350-4358.

26. Leach, A. and Gillet, V., *An Introduction to Chemoinformatics*2007, Dordreecht: Springer. 255.

27. Holliday, J.D. and C-Y, W.P.H., *Grouping of Coefficients for the Calculation of Inter-Molecular Similarity and Dissimilarity using 2D Fragment Bit-Strings.* Combinatorial Chemistry & High Throughput Screening, 2002. **5**(2): p. 155-166.

28. Ballester, P.J., Finn, P.W., and Richards, W.G., *Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology.* Journal of Molecular Graphics and Modelling, 2009. **27**(7): p. 836-845.

29. Kalidas, Y. and Chandra, N., *PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins.* Journal of Structural Biology, 2008. **161**(1): p. 31-42.

30. Laskowski, R., *SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions.* Journal of Molecular Graphics, 1995. **13**(5): p. 323-330.

31. Soga, S., Shirai, H., Kobori, M., and Hirayama, N., *Use of Amino Acid Composition to Predict Ligand-Binding Sites.* Journal of Chemical Information and Modeling, 2007. **47**(2): p. 400-406.

32. Mehio, W., Kemp, G.J.L., Taylor, P., and Walkinshaw, M.D., *Identification of protein binding surfaces using surface triplet propensities*, in *Bioinformatics*2010. p. 2549-2555.

33. Flanagan, C.A., Fromme, B.J., Davidson, J.S., and Millar, R.P., *A High Affinity Gonadotropin-Releasing Hormone (GnRH) Tracer, Radioiodinated at Position 6, Facilitates Analysis of Mutant GnRH Receptors.* Endocrinology, 1998. **139**(10): p. 4115-4119.

Chapter 3. Building a model of the GPCR Gonadotrpin Releasing Hormone Receptor (GnRH-R).

3.1 G-Protein Coupled Receptors as drug targets

The GPCR proteins are a critical target for modern drug development. They are divided into five different families and account for around 2% of the total human genome[1]. Despite 50% of modern drugs being thought to act on GPCRs[2] they are still a relatively poorly understood receptor class due to the difficulties in obtaining high resolution crystal structure data [3]. To date there are still only six crystal structures of GPCRs reported in the literature. The first structure to be solved in 2000 was Bovine Rhodopsin[4]. The publication of the solved structure was a result of the ability to obtain milligram quantities of purified protein relatively easily. This was a landmark paper in protein science and the seven transmembrane helix structure (Figure 2.1) became a foundation for GPCR modelling work for many years[5].

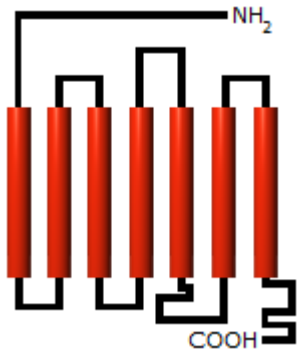


Figure 3-1 Overall topology of the GPCR rhodopsin family. Members of the rhodopsin family possess a conserved structure consisting of 7 Transmembrane α helices (7-TM domain) connected by alternating hydrophilic intracellular and extracellular loops, an amino-terminus on the extracellular side of the membrane and a carboxy-terminus on the intracellular side.

3.1.1 GPCR modelling using rhodopsin as a template structure

As the first crystal structure of any GPCR, rhodopsin was the only template available to model the conformations of other GPCR proteins. Whilst this provided limited structural variation, it did provide modellers with a crucial starting point.

A number of models were created using bovine rhodopsin as the template structure[6-10]; whilst small molecule ligands were successfully found using this approach it does present limitations. Literature studies have evaluated the usefulness of using rhodopsin as a structural template in a variety of GPCR structures. These showed that whilst rhodopsin template modelling was useful in terms of generating a general placement of the 7 transmembrane helices, problems with active site residue placement caused incorrect natural ligand docking[11, 12]. This is a critical problem in ligand discovery[13].

A reason why this occurs is because as the sequence similarity between template and target sequence decreases below 30%, the model generated has a structure which is a reflection of the template more than the target primary sequence[14]. This idea is explored further in Chapter 5.7 where binding of template structure ligands to the target model structure is assessed.

3.1.2 Subsequent crystallisation of rhodopsin-like GPCR proteins offered greater insight to GPCR structure

It was not until seven years later that a new GPCR crystal structure was solved; the human β2-adrenergic receptor was the first ligand activated receptor structure[15, 16]. With this new structural insight into the mechanics of ligand induced movement the understanding of GPCR receptor signalling increased substantially. Over the course

of the next four years there was a relative explosion in the field due in part to production of lysozyme chimeric proteins and a further five structures were published: β1 adrenergic receptor[17, 18], $A2_A$ adenosine receptor[19, 20], CXCR4 chemokine[21], dopamine D3[22] and finally the histamine H_1 receptor[23]. With six crystal structures now available (Table 3-1), significant GPCR insight can be gained. This is particularly true for GPCR activation.

Table 3-1. Summary of current GPCR crystal structures published in the literature.

Receptor	PDB ID	Function	Therapeutic relevance
$A2_A$ Adenosine	3EML	Adenylate cyclase activation[24]	Parkinson's disease[25]
β1 Adrenergic	2VT4	Cardiac regulation[26]	Angina, Glaucoma, Heart disease[27-29]
β2 Adrenergic	2R4R	Cardiac regulation, Muscle action, Digestion[26]	Angina, Heart disease[27, 29]
CXCR4 Chemokine	3ODU	Vascularisation factor[30]	HIV treatment[31]
Dopamine D3	3PBL	Adenylyl cyclase inhibition[32]	Parkinson's disease, ADD, Schizophrenia[32]
Histamine H_1	3RZE	Inflammatory response[33]	Allergy management[33]
Rhodopsin	1F88	Photoreceptor function[34]	-

3.1.3 GPCR activation is a multi-state structural event

The current picture of GPCR activation depends on the structures movement between two key states, the inactive (R) state and the active (R*) state. In the R state the GPCR is still able to bind to and signal through G proteins, however this signalling activity is much lower than is induced upon agonist binding. In the R* state agonist binding causes conformational changes within the GPCR which increase signalling throughput significantly over basal activity. However there is experimental evidence to show that intermediary states exist between these. Fluorescence work on the β2 adrenergic receptor has shown that whilst the ligand free receptor has a preferred conformation, the structure is dynamic (Figure 3-2) and oscillates around this[35]. Agonist binding stabilises the active state but is not strictly required for the formation of R*. Whilst the majority of receptors will be in the R state spontaneous formation of R* is possible such that basal activity has been observed in the absence of agonist molecules[36].

The most significant structural change during activation is the 5-6Å migration of intracellular TM helices 5 and 6 away from the centre of the receptor[37], opening up the space for G protein association. Given the way G proteins are shared among receptors and the way in which one GPCR can signal through multiple pathways it has been revealed that the specificity of G protein association is actually a ligand specific rather than receptor event[38] given that the type of ligand that associates with the GPCR will affect the GPCR conformation in a specific way which will favour a certain G protein to associate.

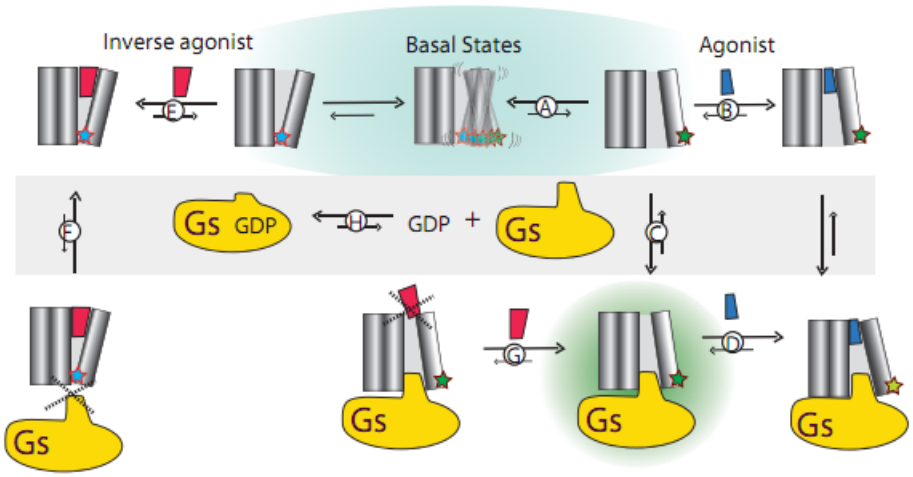


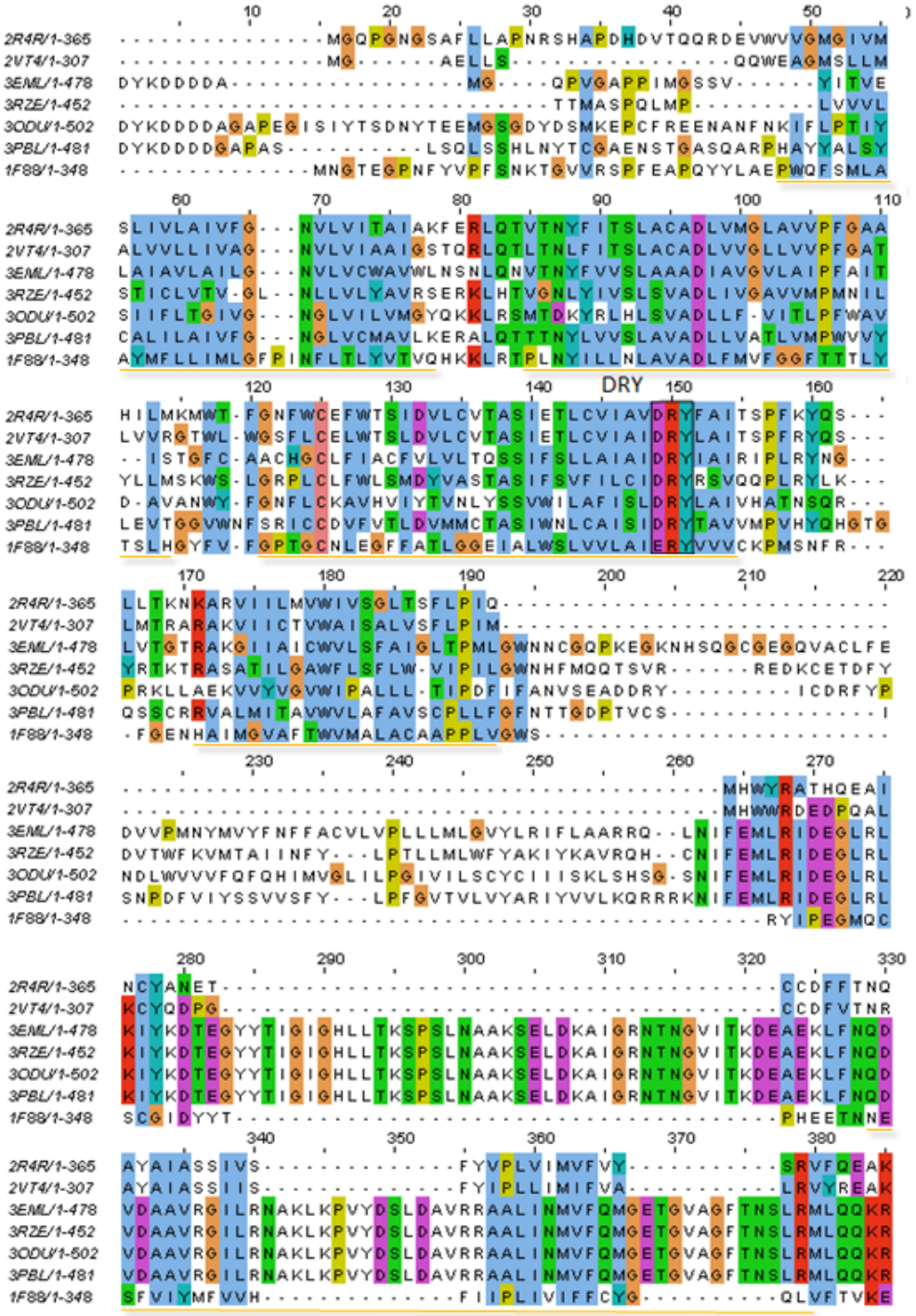
Figure 3-2. Structural conformation changes in the activation of the β2 Adrenergic receptor. Binding of an inverse agonist rather than just an antagonist is the most effective way to prevent the activation of signalling through the G proteins as it prevents association with G_S. Figure taken from reference [39].

3.2 Alignment of GPCR sequences shows low homology outwith key highly conserved domains

Sequence alignments of the receptors (Figure 3-3 and Table 3-2) are surprising in that they have such low sequence identities. For the six proteins in Table 3-1 sequence identities range from 15% to 59% even though they belong to the same class of receptor. The highly conserved glutamic acid/aspartic acid–arginine–tyrosine (E/DRY) domain (Figure 3-3) is a characteristic of the Class A GPCR family, however the significance is still somewhat ambiguous. Initially it was hypothesised by site directed mutagenesis that the motif was essential in maintaining the ground (ligand free) state conformation as a consequence of salt bridge formation between transmembrane (TM) helices 3 and 6. Later mutation experiments of the motif caused constitutive activation (CA) of the receptor[40]. By contrast however, the literature shows that this CA phenomenon is not observed in all members of the

family. A recent review summarised eight GPCRs which did not undergo CA as a result of similar mutations, GnRH-R is amongst those listed[41].

The NPxxY motif (Figure 3-3) is another conserved motif in the aligned sequences and is found on the cytosolic region of the seventh TM helix.



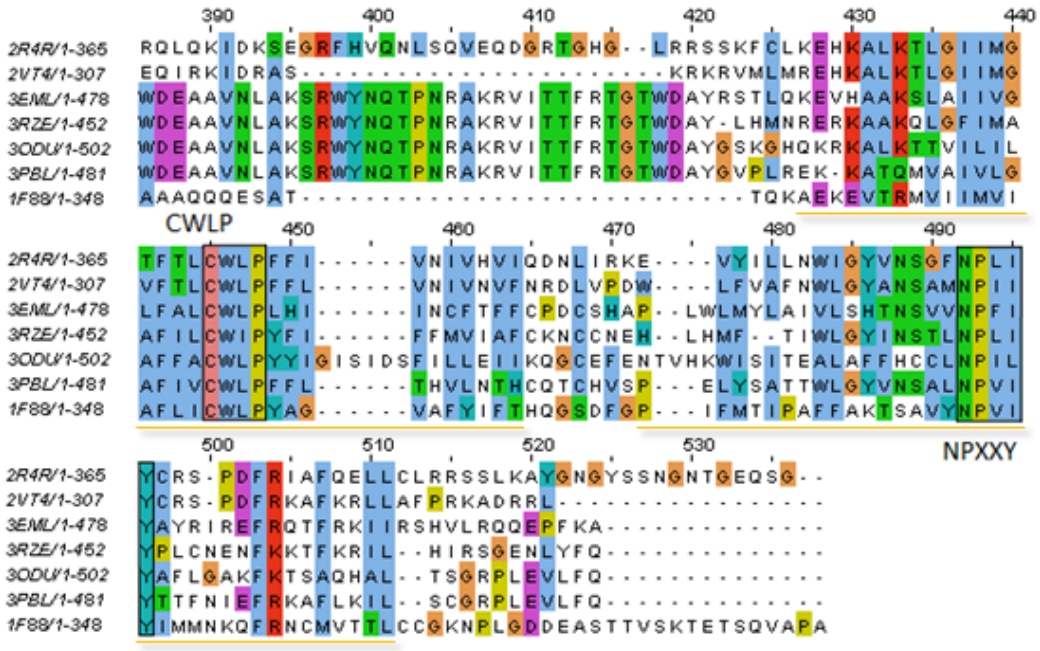


Figure 3-3. Sequence homology of the six GPCR crystal structures. The alignment shows that whilst there is reasonable identity between all GPCR structures the majority of the identity is restricted to the 7-TM helix regions (underscored in orange), most notably the Class A conserved E/DRY motif at the 150 residue region, CWxP around 450 residues in and the NPxxY motif in the 490 residue region.

Table 3-2. Pairwise alignment scores of GPCR crystal structures in a 7 x 7 matrix.

PDB ID	Length	PDB ID	Length	Identity
1F88	348	2R4R	365	15.0
1F88	348	2VT4	307	17.0
1F88	348	3EML	482	16.0
1F88	348	3ODU	502	15.0
1F88	348	3PBL	481	20.0
1F88	348	3RZE	452	14.0

PDB ID	Length	PDB ID	Length	Identity
2R4R	365	2VT4	307	59.0
2R4R	365	3EML	482	24.0
2R4R	365	3ODU	502	17.0
2R4R	365	3PBL	481	28.0
2R4R	365	3RZE	452	25.0
2VT4	307	3EML	482	30.0
2VT4	307	3ODU	502	21.0
2VT4	307	3PBL	481	34.0
2VT4	307	3RZE	452	30.0
3EML	482	3ODU	502	45.0
3EML	482	3PBL	481	53.0
3EML	482	3RZE	452	54.0
3ODU	502	3PBL	481	50.0
3ODU	502	3RZE	452	49.0
3PBL	481	3RZE	452	54.0

3.2.1 Microdomains and switch regions are conserved sequences which apply structural constraints to maintain inactive conformations

As previously noted the E/DRY and NPxxY microdomains are observed to play roles as structural switches or so called “ionic locks” which are responsible for stabilising the R state structure. The Cys-Trp-Xaa-Pro (CWxP) motif is also involved in a similar role, a summary diagram is shown in Figure 3-4.

The E/DRY domain on TM3 (Figure 3-4a) forms one of these ionic locks between a glutamate residue on TM6 and the Arg in the motif, but only when a pivotal asparagine in TM2 is in a protonated state[42].

The tryptophan in the CWxP motif undergoes a close packing interaction with Gly315 in the β2 adrenergic receptor (Figure 3-4c). When the glycine was mutated to an alanine the receptor exhibited constitutive activity demonstrating that the intermolecular force between them locks the receptor in the R state[43]. The R state is a structural conformation

The highly conserved proline residue that is in the NPxxY motif causes a kink in the helix which forces the tyrosine into a pocket formed between TM2, 3, 6 and 7 (Figure 3-4b). Water molecules in this region contribute hydrogen bonding stability to the receptor[44]. It is also a possibility that it may interact with an asparagine (Asn73) on the opposite helix[45].

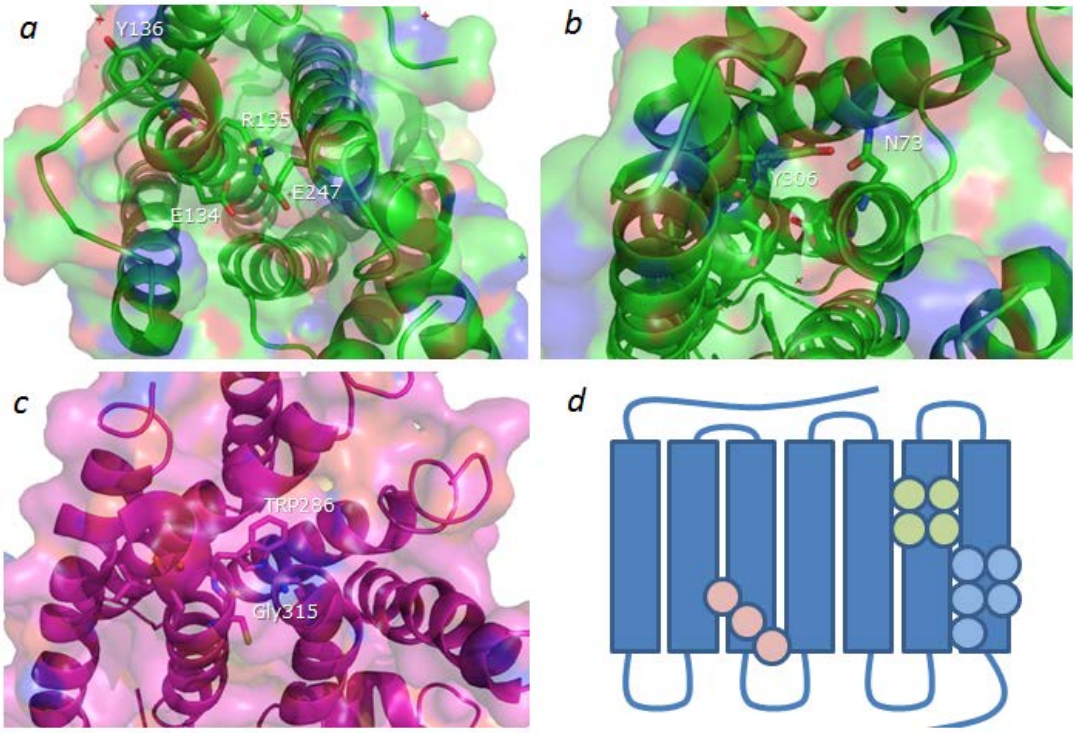


Figure 3-4. Analysis of ionic locks and microdomains in GPCR proteins. (**a**) The E/DRY motif interaction with E247 in rhodopsin. (**b**) The NPxxY motif interaction with N73 also in rhodopsin. (**c**) The CWxP motif interaction as observed in the β2-adrenergic receptor. (**d**) A cartoon schematic of the GPCR structure depicting where the DRY motif (Pink), CWxP motif (Green) and NPxxY (light blue) motifs are located.

3.3 Alignment of GPCR crystal structures highlights favourable overlap of seven transmembrane helices.

Alignment of the receptors by 3D structure rather than sequence is more illustrative of how the receptors share identity (Figure 3-5). Using rhodopsin as the fixed structure and aligning the other GPCRs against it, it can be seen that whilst sequence identity is low the structures themselves superpose well. Taking rhodopsin and dopamine D3 receptor as an example, even though sequence identity score is relatively low (20%) the structures themselves superpose better than might be expected (RMSD = 2.7Å). The pairwise alignment was performed using 197 Cα atoms from each structure using PyMol.

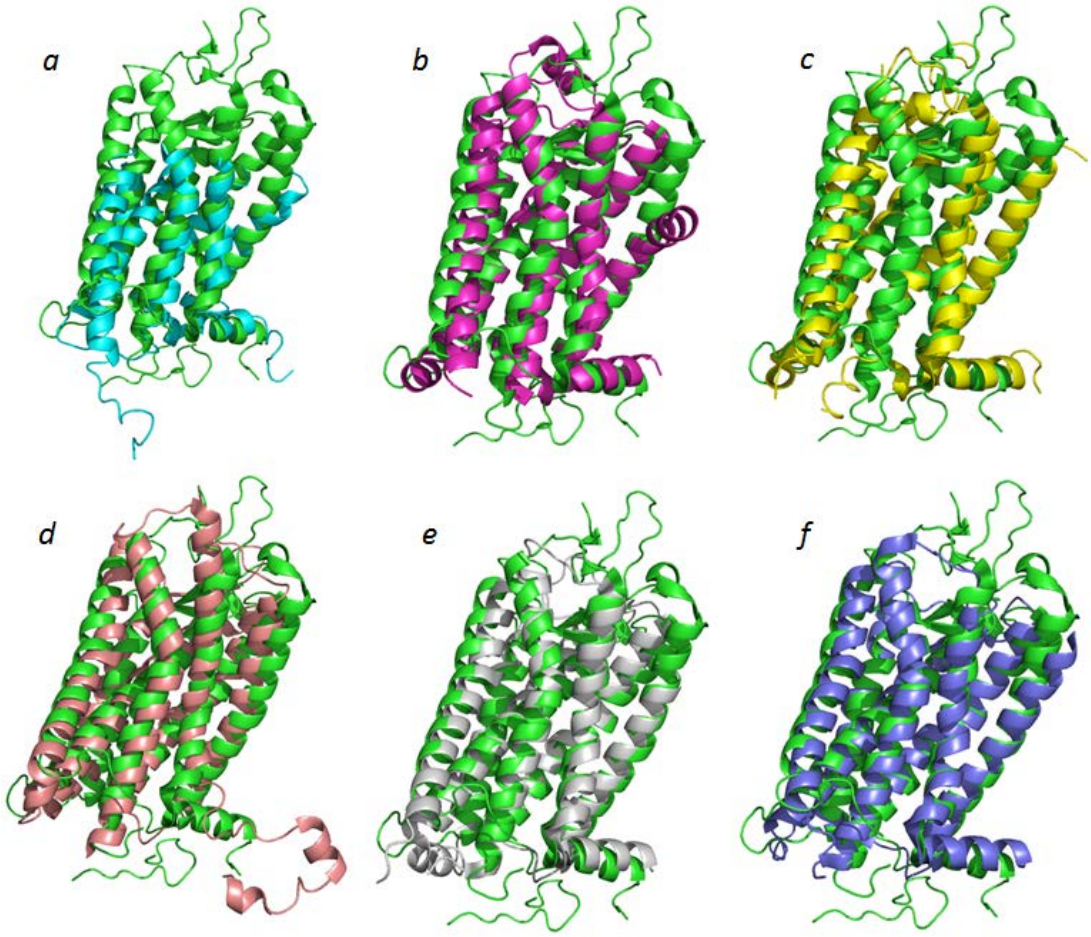


Figure 3-5. Structural alignment of GPCRs using rhodopsin (green) as the anchored protein. (a) Rhodopsin alignment with the β2 adrenergic receptor, Root Mean Square Distance (RMSD) = 4.1Å (b) Rhodopsin alignment with the β1 adrenergic receptor, RMSD = 3.3Å (c) Rhodopsin alignment with A2A adenosine receptor, RMSD = 3.0Å (d) Rhodopsin alignment with CXCR4 chemokine receptor, RMSD = 4.6Å (e) Rhodopsin alignment with dopamine D3 receptor, RMSD = 2.7Å (f) Rhodopsin alignment with histamine H1 receptor, RMSD = 3.4Å.

3.4 Discussion of sequence and structure alignments of the GPCR proteins

Alignment results showed that there are only 13 residues which are fully conserved across all six of the GPCR sequences. Given that these proteins are 300-500 amino acids long this is a remarkably low conservation. This observation has been previously commented on in regard to evolutionarily “ancient” genes, where gene duplication and subsequent mutation allows a gene to change function whilst free of selection pressure[46].

The relatively low sequence identity between the GPCRs could be explained as a consequence of successful mutations after a gene duplication event from a common ancestor giving rise to the large number of GPCRs that have been found with a wide range of signalling roles[47].

3.5 Building a homology model of GnRH-R using a rhodopsin template

As observed previously there is a low identity between residue type in GPCR proteins except in the case of structurally conserved microdomains. Therefore there is a difficulty in modelling an accurate receptor structure from one homology model. This is why in this work both rhodopsin and the β2-adrenergic receptor structures were used to create models.

Three different model structures were made; the first two models were derived using homology modelling with the rhodopsin structure and were modified to reflect the expected positions of the receptor in the active and inactive states. The final model was created using the β2-adrenergic receptor structure and was made to reflect the inactive state.

3.5.1 Rhodopsin and GnRH-R share low sequence homology

Alignment of GnRH-R and rhodopsin showed an identity score of 15% (Figure 3-6); whilst it would have been preferable to have a larger identity to make our model more reliable further refinement should produce a usable model structure[48].

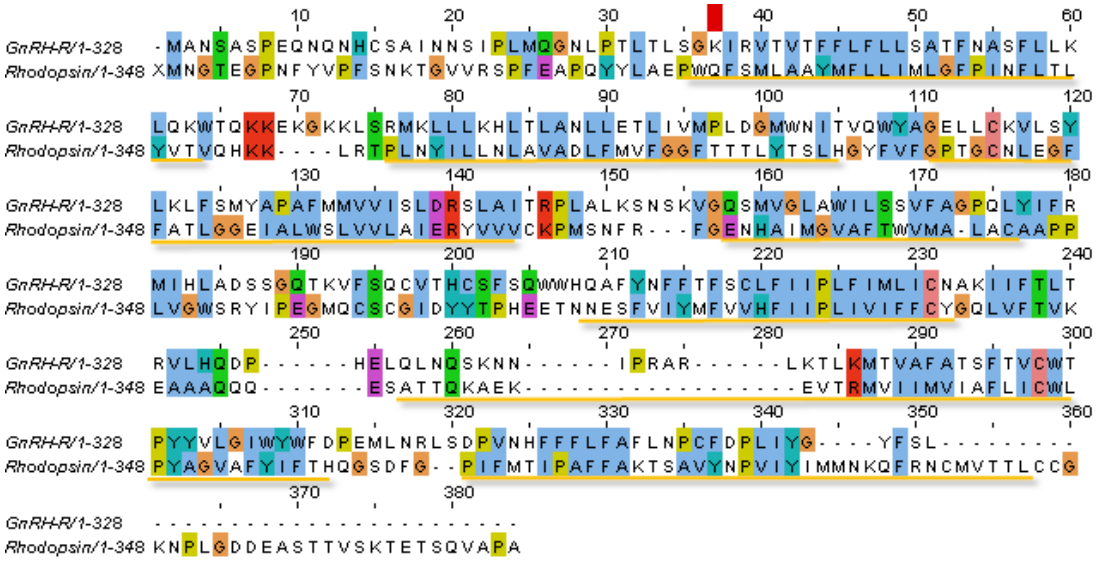


Figure 3-6. Identity scoring of GnRH-R and rhodopsin residues. When aligning the two GPCRs the identity was only 12%. It is interesting that GnRH-R has substitutions in a number of the key ionic lock microdomains. The DRY motif is a DRS and the NPxxY motif is DPxxY. Transmembrane helices are underscored in orange.

3.5.2 Homology modelling of the GnRH receptor based on rhodopsin

In order to generate a homology model using MODELLER[49] the input PDB file 1U19 for the crystal structure of rhodopsin was used as it is currently the highest resolution structure in the PDB. After input of the GnRH-R I sequence into MODELLER five homology structures were generated and compared through superposition (Figure 3-7). The model quality was then assessed using ProSA-web[50] (Figure 3-8). This is a web server which evaluates distance pair potentials of the Cα backbone in the model as well as solvent exposure of each residue[51-53]. The *z* score generated for the structure indicates how the structure compares against protein scores computed for the entire PDB and calculates the model quality by measuring deviation of total energy in the structure against the energy distribution of the model in random conformations. This is then plotted alongside scores from proteins of similar size to allow comparison.

Alongside the GnRH-R model created from rhodopsin, the rhodopsin crystal structure was also evaluated with ProSA-web tool and a *z* score was obtained. This allows us to compare the *z* scores for the GnRH-R model and the rhodopsin crystal structure.

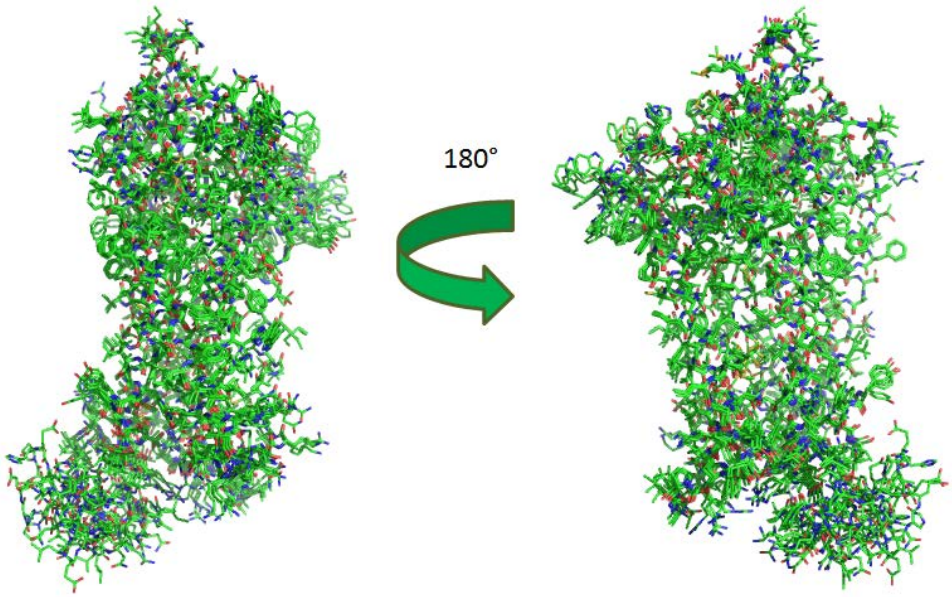


Figure 3-7. Superposing the five GnRH-R model structures generated by MODELLER. Superposing the models showed how well the program was able to assess the positions of the flexible loop regions by fitting all structures to "best" model and looking at the variance in RMSD value. The RMSD value range was 0.23-0.38Å across the models and it can be seen that the TM regions are well defined whilst the IC and EC loops are not.

In general the model scored fairly well. Figure 3-8 shows that the *z* score is in line with other proteins of similar size albeit in the outlying regions of the mean energy plot for x-ray structural data. Similarly the regions of high energy were localised to the portions of the helices which maintain high flexibility. Whilst the rhodopsin crystal structure produced a better *z* score, it too is placed near the edge of the x-ray results for similar size proteins. This suggests that the software may have greater difficulty analysing membrane bound proteins.

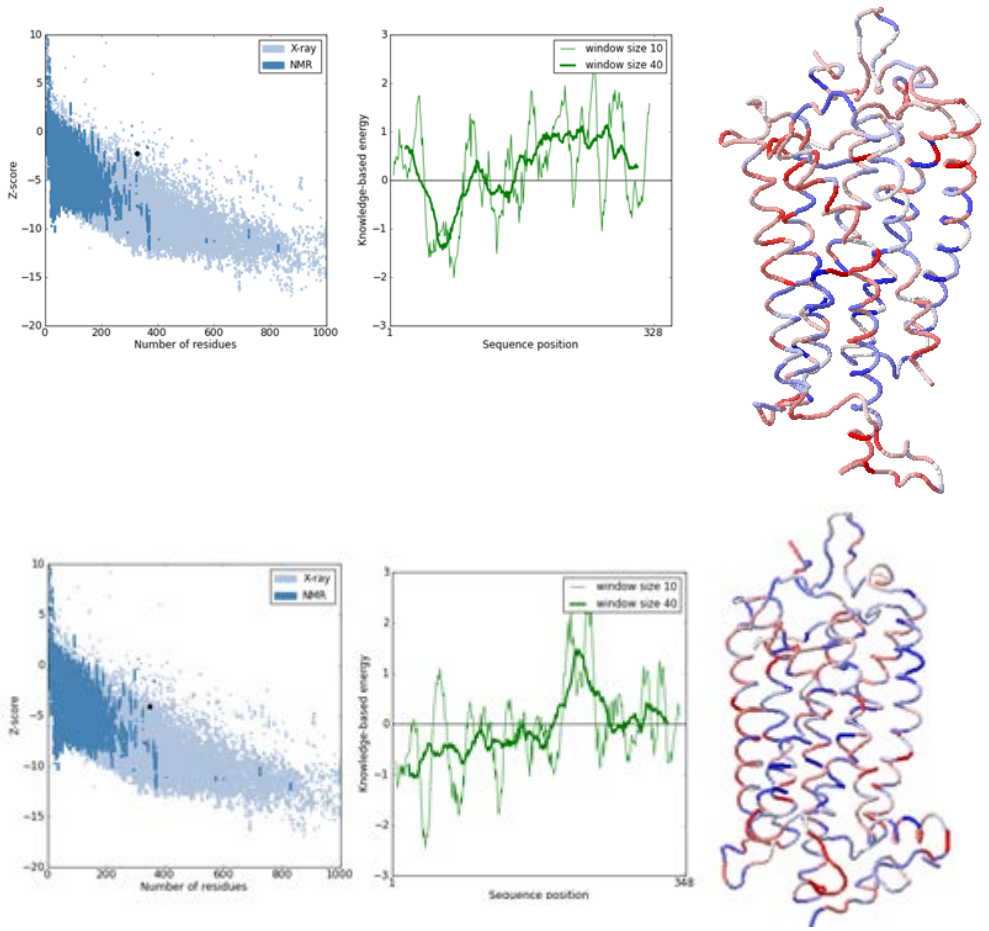


Figure 3-8. Using ProSA Web as a tool to check model quality. The best scoring GnRH-R model (Top) had a *z* score of -2.24. When this is plotted against proteins of a similar size the score is on the outlying regions of that obtained from X-ray crystallography but is still a reasonable fit (Top Left, black dot shows GnRH-R model score). Looking at energy potential of individual residues shows 22 residues which have energy values out with the norm (Top Middle and Top Right). The crystal structure of rhodopsin was also evaluated (Bottom) and a *z* score of -4.11was calculated.

3.5.3 Evaluation of GnRH-R rhodopsin-based homology models

Next the differences between the models in terms of active site structural accuracy were examined according to the site-mutagenesis and peptide/non-peptide binding data[54-56] (Figure 3-9). The positions of the key GnRH peptide binding residues in the active and inactive state, as derived experimentally, were compared to the rhodopsin based GnRH-R model. Theoretically if the modelled positions of the key amino acids

are in close positions to those experimentally derived then there is higher confidence in the model quality.

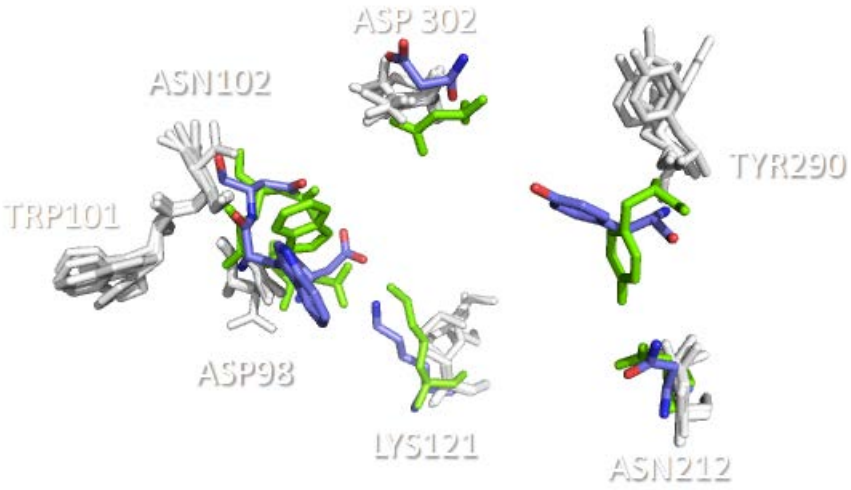


Figure 3-9. Evaluation of binding residue placement in MODELLER generated structure from rhodopsin. The five model structures are shown in white, residues in the experimentally determined ligand bound structure are shown in green, the ligand unbound state is in blue.

Comparison between the active and inactive states shows that the model was able to predict Lys121, Asn212 and Asp302 positions fairly closely to the expected positions. The Asp98, Trp101 and Asn102 cluster along with Tyr290 are displaced from the expected position by the rotation on their respective helices.

3.5.4 Flexible intracellular and extracellular loop regions present challenges in the model building process

When docking experiments are carried out most of the extracellular loop regions will have to be removed as they will interfere with the docking into the active site. As loop refinement is a time consuming stage in model creation it may initially appear beneficial to remove them from the model entirely. However the positioning of the loops on both the intracellular and extracellular sides play a crucial role in defining

the rotation of the helices in the model, and are preserved and modelled as closely as possible given the information known.

3.5.4.1 Loop optimisation

Before adjusting the necessary active site residues manually the IC and EC loop regions were further optimised using the loop refinement module of MODELLER. The portions of the structure which had been highlighted by the ProSA Web energy scoring function were defined as the target regions for additional refinement. These regions were 96-110, 169-181, 211-223, 236-259 and 275-294 (Figure 3-7). After several cycles model ligand binding site positions were in close agreement with the experimental results and optimisation was deemed to be complete. This was particularly important because of the disulphide bonds between Cys14→Cys200 and Cys114→Cys195 on the ligand binding face[57].

3.5.5 Preserving ionic locks and microswitch domain interactions

As discussed the GPCR ionic locks and microswitch domains are an important part of receptor function. Therefore in order to have the most representative receptor model it is useful to ensure that these are spatially accurate. There is however a degree of freedom with regards to how well the motifs fit into the model as they are are key players in structure activation and the method of activation between GnRH-R and rhodopsin differs, in that rhodopsin is not ligand activated whilst GnRH-R is.

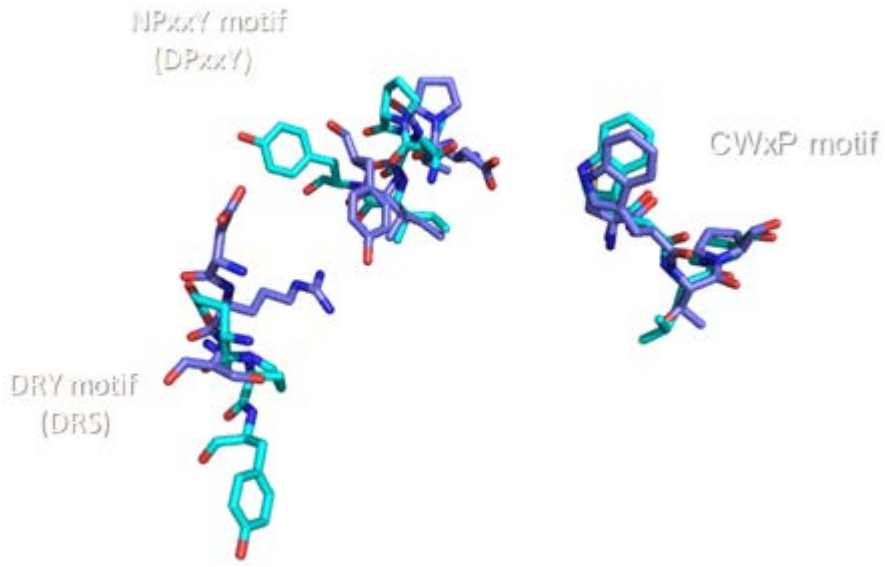


Figure 3-10. Comparison of microdomains between rhodopsin derived GnRH-R model (dark blue) and rhodopsin (cyan). The DRY ,CWxP and NPxxY motifs appear to align fairly well between both models, there are small deviations which are minor translation/rotational differences. Motifs which differ from the canonical residues are shown in parentheses.

Our superposition of the three motifs (Figure 3-10) has shown that they align well; the CWxP motif in particular is very close (equivalent C_α- C_α distances are less than 2Å). The DRS motif differs the most, a possible explanation for this may be due to DRY motif being a critical regulator of receptor ground state (Characterization of Intracellular Signaling Mediated by Human Somatostatin Receptor 5: Role of the DRY Motif and the Third Intracellular Loop) As rhodopsin is covalently bound to the ligand the ground state activation method will differ.

3.6 GnRH-R modelling using the β2-adrenergic receptor

As previously discussed the β2-adrenergic receptor was the first ligand activated GPCR crystal structure, making it a more appropriate template structure for a GnRH-R homology model though neither rhodopsin nor the β2-adrenergic receptor have sequence identities higher than 15% (Figure 3-6). The β2-adrenergic receptor does however have a higher sequence identity to GnRH-R than rhodopsin (Figure 3-11)

3.6.1 Structure generation and loop optimisation of the GnRH-R model based on the β2-adrenergic receptor

As in the previous GnRH-R modelling MODELLER was used to build the homology model. The PDB file 2R4R was used as the template structure file and the models produced were scored using ProSA Web (Figure 3-12).

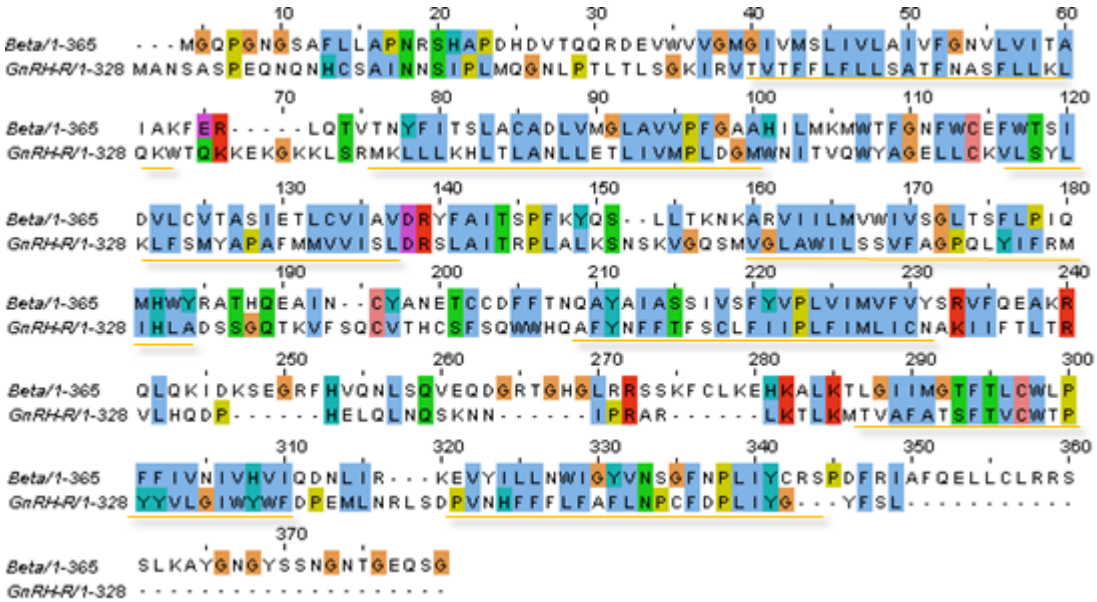


Figure 3-11. Alignment of GnRH-R and the β2-adrenergic receptor. The identity score for the alignment was 15% which is slightly higher than the 13% obtained from GnRH-R and rhodopsin. Since the β2-adrenergic receptor is also a ligand activated GPCR it is therefore a more suitable structure for homology modelling. Transmembrane helices are underscored in orange.

Additionally, the crystal structure of the β2-adrenergic receptor was also scored using ProSA-Web so that the *z* score could be compared between the two structures. The GnRH-R model created using the β2-adrenergic receptor structure had a calculated *z* score of -5.42. The β2-adrenergic receptor structure itself had a *z* score of -7.01 (Figure 3-12).

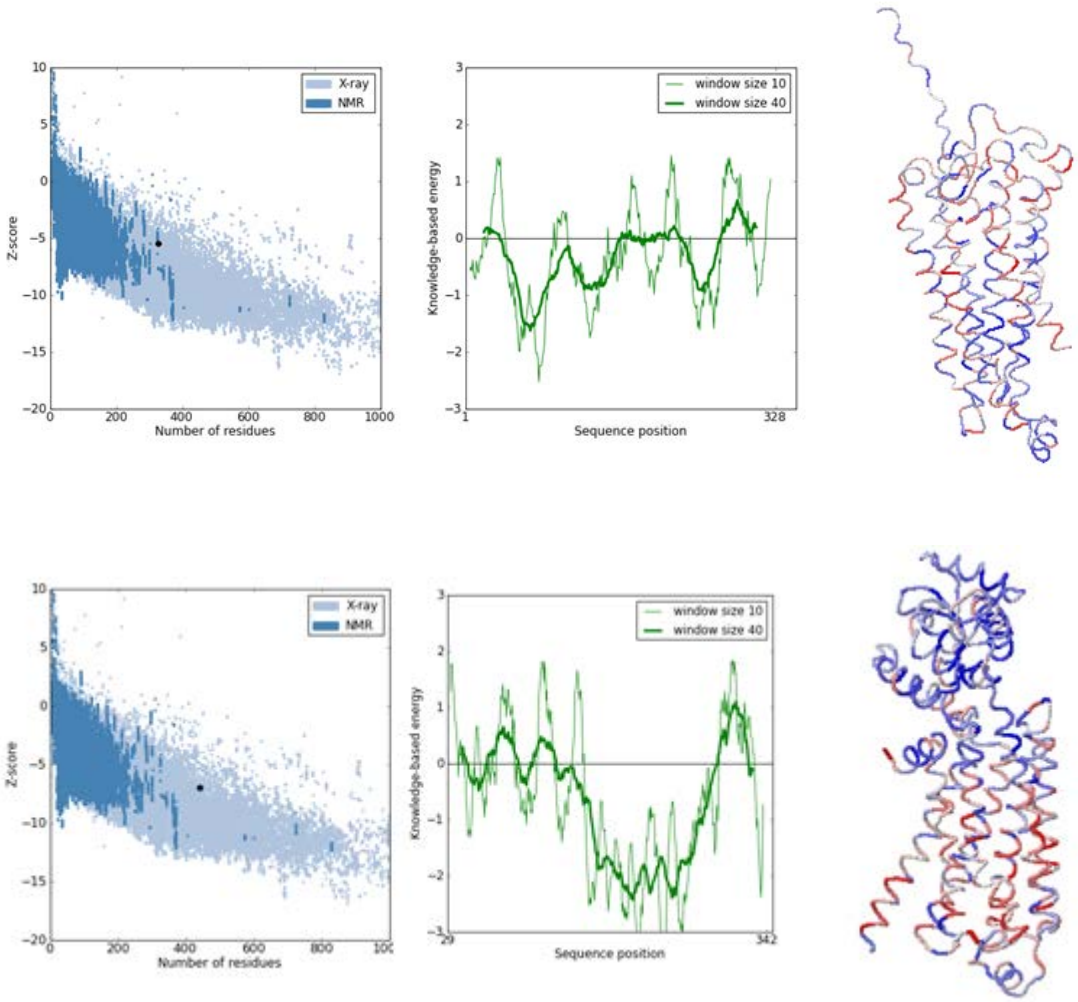


Figure 3-12. ProSA-Web analysis of the GnRH-R β2-adrenergic receptor homology model (Top) and β2-adrenergic receptor structure (Bottom) . This model scored more favourably than the GnRH-R model from rhodopsin (*z* score -5.42). However the first flexible loop region is quite poorly refined. Similarly the β2-adrenergic receptor structure had a better *z* score compared to the rhodopsin structure.

As the first EC loop region was highly disordered after initial modelling an attempt was made to refine this region further using MODELLER. However the results were unnsatisfactory, and instead that part of the structure was rebuilt using the rhodopsin derived model part as the template and performing energy minimisation steps to tighten the structure.

It is an interesting observation that the proposed model quality from the β2-adrenergic receptor template scores better than that from rhodopsin. The *z* score is much closer to other proteins of its size than was seen with the rhodopsin derived model. Furthermore the β2-adrenergic receptor structure itself also scores more favourably; this indicates that the program does not have problems with the GPCR proteins as a whole, though rhodopsin may present a unique challenge.

3.7 A comparison between models based on rhodopsin and the β2-adrenergic receptor

3.7.1 Superposing the ligand free homology models of GnRH-R from rhodopsin and the β2-adrenergic receptor

Now that a model has been created from a receptor with a constitutively bound ligand and one from a ligand activated receptor, it is interesting to look at the structural differences/similarities between them (Figure 3-13).

The most obvious difference from the alignment is the position of TM6. Because the rhodopsin structure used as a template contained the ligand molecule this helix is in the active conformation which involves a 5-6Å movement of this helix in the distal plane. This can be checked through alignment of the ligand free opsin structure with

rhodopsin (RMSD = 1.6Å), which does indeed show that the opsin TM6 adopts a similar orientation to that of the β2-adrenergic receptor model.

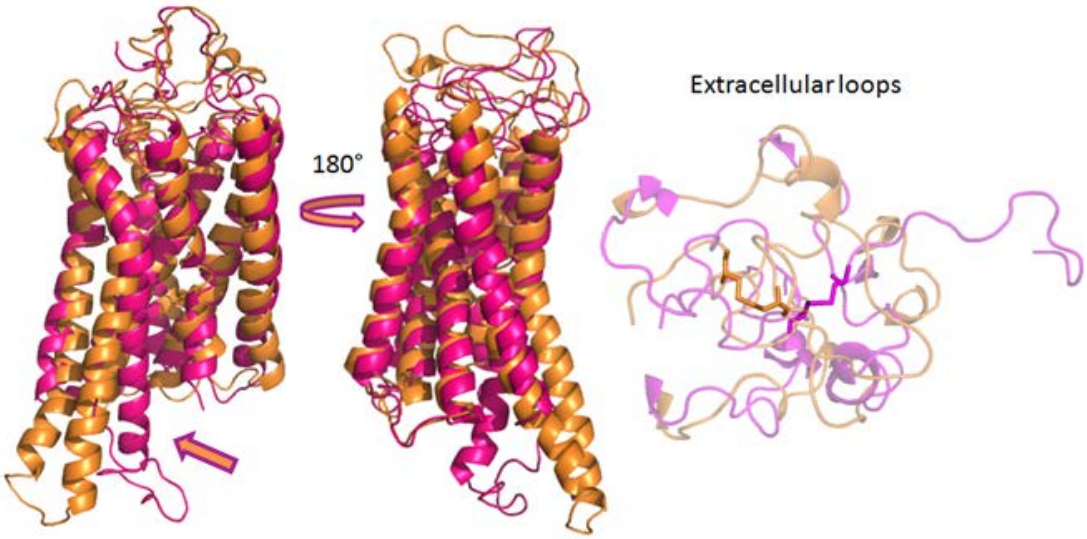


Figure 3-13. Structural alignment of GnRH-R models created using rhodopsin (magenta) and the β2-adrenergic receptor (orange). The two structures align with an RMSD value of 3.7Å which is a typical value for GPCR structure alignments and highlights a deviation in positioning of the TM6 helix (highlighted by arrow). C14 and C200 disulphide bonding is shown in sticks in the extracellular loop.

Relative positioning of the disulphide bonds between Cys14→Cys200 is also slightly different. There is a distance of 6.1Å between the Cα of the Cys14 residues and 9.2Å between the Cα of the Cys200 residues. However disulphide bonds remain preserved. This should not be detrimental to small molecule docking as there is no known intramolecular interaction involving these residues and the target binding site is more deeply buried into the centre of the pocket.

3.7.2 Examining differences in positioning of the conserved residues between model structures

Examining the positions of the key conserved residues between both models and their respective templates, we observe variation in residue orientation and spacial

positioning (Figure 3-14). This is unexpected as the most heavily conserved residues are typically functionally important and should therefore occupy similar spaces. However it is possible that the difference is a result of conformational changes induced as a result of specific changes in structure to accommodate respective ligands. This highlights the difficulties of modelling protein structures with so few available crystal structures. As no consensus can be reached on how the helices will shift to accommodate ligand binding there is difficulty in determining which model is likely to be the most accurate.

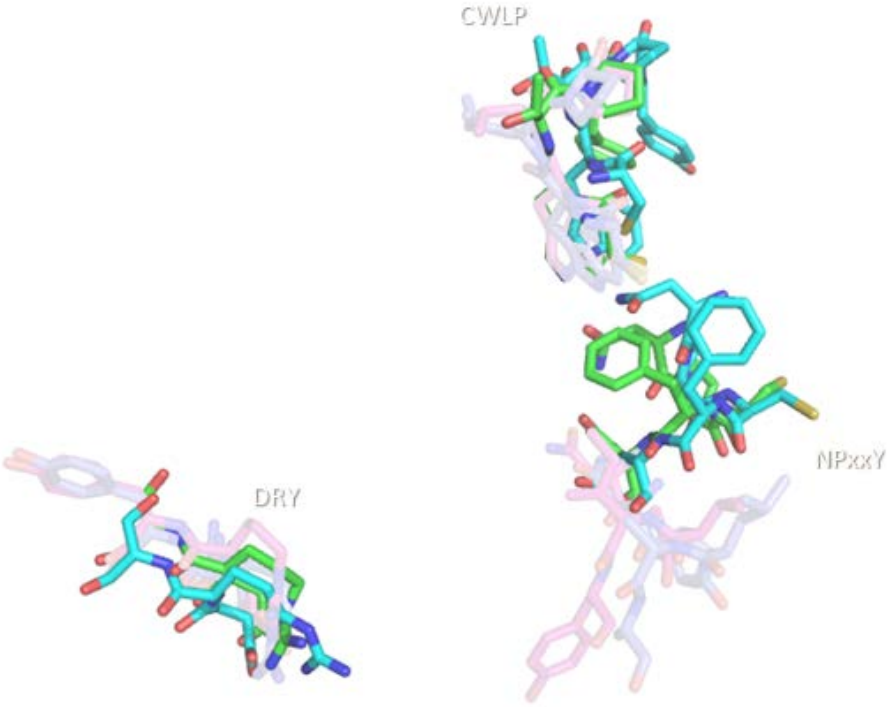


Figure 3-14. Comparison of key conserved residues positions between the rhodopsin (green) and the β2-adrenergic receptor (cyan) derived homology models. Positions of the same residues in the rhodopsin (transparent purple) and β2-adrenergic receptor (transparent dark blue) crystal structure are also shown.

3.8 Understanding GPCR activation and signalling

The signal transduction model of the receptor was first postulated in the Ternary Complex Model (TCM)[58]. This model proposed that GPCR activation occurred only

when agonist, receptor and G protein formed a complex in an agonist induced interaction. This was modified after it was found that ligand free GPCR proteins could change into an active state and bind to G-proteins[59]. As a result of the increase in available structural data, understanding of how the proteins are able to transmit signalling information across the plasma membrane has increased dramatically.

3.8.1 Insights into GnRH-R signalling from the newly crystallised β2-adrenergic receptor-Gs protein complex

The first crystal structure of a GPCR-G protein complex has recently been published in the literature[60] by the same group who published the β2-adrenergic receptor structure. Again the receptor was stabilised using a T4 lysosyme chimera protein and bound to a high affinity agonist. The structure (PDB ID 3SN6) was aligned against our GnRH-R (β2-adrenergic receptor) homology model and we looked at the model-G-protein complex interface to assess possible polar contacts within a 3.2Å distance between the GnRH receptor and the Gs protein. The interactions found were then compared with those of the β2-adrenergic receptor and the Gs protein.

In the β2AR-Gs complex 20 β2AR residues are involved in creating 35 interactions to the Gs protein. Of the 20 unique residues 7 of them are also conserved within GnRH-R, these are residues R131, A134, I135, P138, S143 and L230. Of these 7 residues only alanine is both conserved and potentially able to interact with the Gs protein. In the GnRH-R model there are 13 unique residues which can potentially form 19 interactions with Gs. These are shown in Table 3-3 and Figure 3-14.

Table 3-3. potential intermolecular interactions between GnRH-R and Gs.

GnRH-R residue	Gs residue	GnRH-R residue	Gs residue	GnRH-R residue	Gs residue
Lys72	Tyr391	Gln244	Asp381 (2)	Asn257	Tyr358
Ala142	His387	Glu248	Arg374	Ala261	Leu394 (2)
Lys150	Lys34, Gln 35	Gln250	Arg342 (2), Asp323 (2)	Thr265	Leu393 (2)
Asn152	Lys34	Gln253	Tyr350 (2), Gly353	Ser254	Arg385
Thr237	Gln384	-	-	-	-

This new crystal structure gives an interesting insight into the possible binding mechanism between GnRH-R and the G protein coupling interactions. Whilst it is noted that Gs is not the natural G protein which couples with GnRH-R, there is a high identity between Gs and $G_{q/11}$ so it is likely that the two proteins have similar interaction mechanisms. This could give rise to the possibility of disrupting the GnRH-R signalling pathway through interruption of direct coupling between GnRH-R and Gs rather than through blocking receptor activation at the extracellular face.

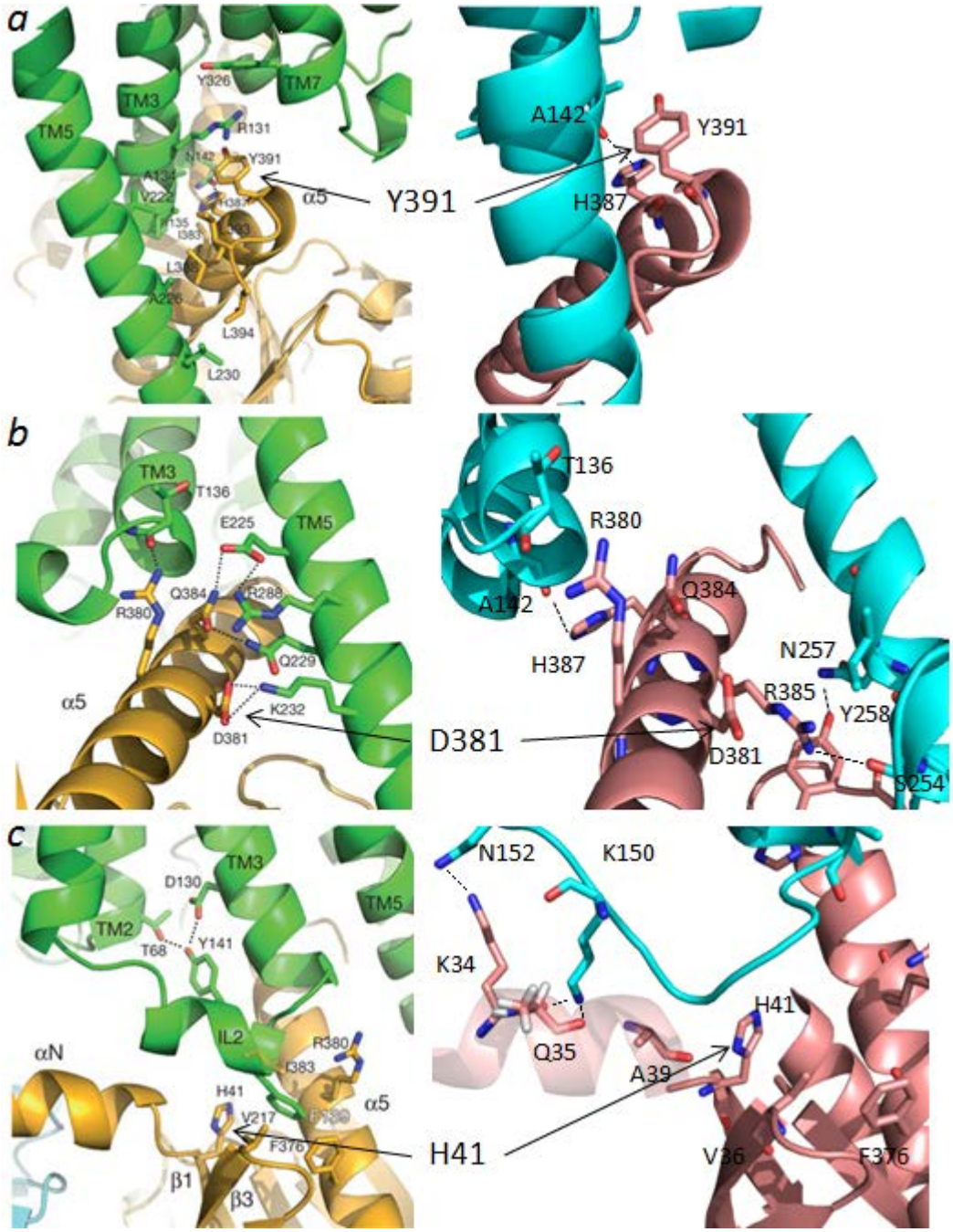


Figure 3-15. Comparison of Gs protein binding between β2-adrenergic receptor (left) and GnRH-R β2-adrenergic receptor homology model (right). Whilst the GnRH receptor is not known to bind to the Gs protein it does bind to $G_{q/11}$ which shares high identity (39%) with Gs.

Investigating the potential interactions that could be targeted between GnRH-R and $G_{q/11}$ opens up the potential for new studies into GPCR drug design which can inhibit

basal activity whilst leaving the extracellular binding domain free for agonist binding. This G protein focussed approach may allow greater control of signalling from G protein coupled receptors which are able to bind to and signal via more than a single type of G protein receptor.

This study will however focus on the targeting of the extracellular GnRH-R binding site for binding of small molecule inhibitors via modelling and validating the extracellular binding site which is discussed in the next chapter.

3.9 References

1. Fredriksson, R., et al., *The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints.* Molecular Pharmacology, 2003. **63**(6): p. 1256-1272.

2. Hancock, A.A., *The challenge of drug discovery of a GPCR target: Analysis of preclinical pharmacology of histamine H3 antagonists/inverse agonists.* Biochemical Pharmacology, 2006. **71**(8): p. 1103-1213.

3. Hubbard, R.E., *Structure-based drug discovery and protein targets in the CNS.* Neuropharmacology. **60**(1): p. 7-23.

4. Palczewski, K., et al., *Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor*, 2000. p. 739-745.

5. Tebben, A. and D. Schnur, *Chemoinformatics And Computational Chemical Biology*. Vol. 672. 2011: Springer. 359-386.

6. Evers, A. and T. Klabunde, *Structure-based Drug Discovery Using GPCR Homology Modeling: Successful Virtual Screening for Antagonists of the Alpha1A Adrenergic Receptor.* Journal of Medicinal Chemistry, 2005. **48**(4): p. 1088-1097.

7. Furse, K.E. and T.P. Lybrand, *Three-Dimensional Models for β-Adrenergic Receptor Complexes with Agonists and Antagonists.* Journal of Medicinal Chemistry, 2003. **46**(21): p. 4450-4462.

8. Kiss, R., Z. Kovári, and G.M. Keserű, *Homology modelling and binding site mapping of the human histamine H1 receptor.* European Journal of Medicinal Chemistry, 2004. **39**(11): p. 959-967.

9. Oliveira, L., et al., *Heavier-than-air flying machines are impossible.* FEBS Letters, 2004. **564**(3): p. 269-273.

10. Varady, J., et al., *Molecular Modeling of the Three-Dimensional Structure of Dopamine 3 (D3) Subtype Receptor: Discovery of Novel and Potent D3 Ligands through a Hybrid Pharmacophore- and Structure-Based Database Searching Approach.* Journal of Medicinal Chemistry, 2003. **46**(21): p. 4377-4392.

11. Axe, F.U., S.D. Bembenek, and S. Szalma, *Three-dimensional models of histamine H3 receptor antagonist complexes and their pharmacophore.* Journal of Molecular Graphics and Modelling, 2006. **24**(6): p. 456-464.

12. Costanzi, S., *On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the β2-adrenergic receptor.* J Med Chem, 2008. **51**(10): p. 2907-2914.

13. Schlyer, S. and R. Horuk, *I want a new drug: G-protein-coupled receptors in drug development.* Drug Discovery Today, 2006. **11**(11–12): p. 481-493.

14. Tramontano, A. and V. Morea, *Assessment of homology-based predictions in CASP5.* Proteins: Structure, Function, and Bioinformatics, 2003. **53**(S6): p. 352-368.

15. Cherezov, V., et al., *High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein-Coupled Receptor.* Science, 2007. **318**(5854): p. 1258-1265.

16. Rasmussen, S.G.F., et al., *Crystal structure of the human β2 adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-487.

17. Warne, T., et al., *Structure of a β1-adrenergic G-protein-coupled receptor.* Nature, 2008. **454**(7203): p. 486-491.

18. Warne, T., et al., *The structural basis for agonist and partial agonist action on a β1-adrenergic receptor.* Nature, 2011. **469**(7329): p. 241-244.

19. Lebon, G., et al., *Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation.* Nature, 2011. **474**(7352): p. 521-525.

20. Jaakola, V.-P., et al., *The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist.* Science, 2008. **322**(5905): p. 1211-1217.

21. Wu, B., et al., *Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists*. p. 1066-1071.

22. Chien, E.Y.T., et al., *Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist.* Science. **330**(6007): p. 1091-1095.

23. Shimamura, T., et al., *Structure of the human histamine H1 receptor complex with doxepin.* Nature. **475**(7354): p. 65-70.

24. Moreau, J.-L. and G. Huber, *Central adenosine A2A receptors: an overview.* Brain Research Reviews, 1999. **31**(1): p. 65-82.

25. Pinna, A., et al., *New therapies for the treatment of Parkinson's disease: Adenosine A2A receptor antagonists.* Life Sciences, 2005. **77**(26): p. 3259-3267.

26. Hein, L. and B.K. Kobilka, *Adrenergic Receptors From Molecular Structure to in vivo function.* Trends in Cardiovascular Medicine, 1997. **7**(5): p. 137-145.

27. McHugh, J., et al., *Beta-blockers in the management of cardiovascular diseases.* Osteopathic Family Physician, 2010. **2**(5): p. 131-138.

28. Grieshaber, M.C. and J. Flammer, *Is the medication used to achieve the target intraocular pressure in glaucoma therapy of relevance? - An exemplary analysis on the basis of two beta-blockers.* Progress in Retinal and Eye Research, 2010. **29**(1): p. 79-93.

29. Basile, J.N., *One Size Does Not Fit All: The Role of Vasodilating [beta]-Blockers in Controlling Hypertension as a Means of Reducing Cardiovascular and Stroke Risk.* The American Journal of Medicine, 2010. **123**(7, Supplement 1): p. S9-S15.

30. Tachibana, K., et al., *The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract.* Nature, 1998. **393**(6685): p. 591-594.

31. Miller, J.F., et al., *Novel N-substituted benzimidazole CXCR4 antagonists as potential anti-HIV agents.* Bioorganic & Medicinal Chemistry Letters, 2010. **20**(7): p. 2125-2128.

32. Maggio, R. and M.J. Millan, *Dopamine D2-D3 receptor heteromers: pharmacological properties and therapeutic significance.* Current Opinion in Pharmacology, 2010. **10**(1): p. 100-107.

33. Simons, F.E.R., *H1-antihistamines: More relevant than ever in the treatment of allergic disorders.* Journal of Allergy and Clinical Immunology, 2003. **112**(4, Supplement 1): p. S42-S52.

34. Hurley, J.B., M. Spencer, and G.A. Niemi, *Rhodopsin phosphorylation and its role in photoreceptor function.* Vision Research, 1998. **38**(10): p. 1341-1352.

35. Ghanouni, P., et al., *Functionally Different Agonists Induce Distinct Conformations in the G Protein Coupling Domain of the Î²2Adrenergic Receptor*, 2001. p. 24433-34436.

36. Naor, Z., *Signaling by G-protein-coupled receptor (GPCR): Studies on the GnRH receptor.* Frontiers in Neuroendocrinology, 2009. **30**(1): p. 10-29.

37. Scheerer, P., et al., *Crystal structure of opsin in its G-protein-interacting conformation.* Nature, 2008. **455**(7212): p. 497-502.

38. Galandrin, S., G. Oligny-Longpré, and M. Bouvier, *The evasive nature of drug efficacy: implications for drug discovery.* Trends in Pharmacological Sciences, 2007. **28**(8): p. 423-530.

39. Yao, X.J., et al., *The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex*, 2009. p. 9501-9506.

40. Zhang, M., et al., *The Formation of a Salt Bridge Between Helices 3 and 6 Is Responsible for the Constitutive Activity and Lack of Hormone Responsiveness of the Naturally Occurring L457R Mutation of the Human Lutropin Receptor.* Journal of Biological Chemistry, 2005. **280**(28): p. 26169-26176.

41. Rovati, G.E., V. Capra, and R.R. Neubig, *The Highly Conserved DRY Motif of Class A G Protein-Coupled Receptors: Beyond the Ground State.* Molecular Pharmacology, 2007. **71**(4): p. 959-964.

42. Vanni, S., et al., *A Conserved Protonation-Induced Switch can Trigger "Ionic-Lock" Formation in Adrenergic Receptors.* Journal of Molecular Biology, 2010. **397**(5): p. 1339-1349.

43. Arakawa, M., et al., *Structural and functional roles of small group-conserved amino acids present on helix-H7 in the [beta]2-adrenergic receptor.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2011. **1808**(4): p. 1170-1178.

44. Rosenbaum, D.M., S.G.F. Rasmussen, and B.K. Kobilka, *The structure and function of G-protein-coupled receptors.* Nature, 2009. **459**(7245): p. 356-363.

45. Fritze, O., et al., *Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation.* Proceedings of the National Academy of Sciences, 2003. **100**(5): p. 2290-2295.

46. Schöneberg, T., et al., *Learning from the past: evolution of GPCR functions.* Trends in Pharmacological Sciences, 2007. **28**(3): p. 117-121.

47. Strotmann, R., et al., *Evolution of GPCR: Change and continuity.* Molecular and Cellular Endocrinology, 2011. **331**(2): p. 170-178.

48. Fanelli, F. and P.G. De Benedetti, *Computational Modeling Approaches to Structure and Function Analysis of G Protein-Coupled Receptors.* Chemical Reviews, 2005. **105**(9): p. 3297-3351.

49. Eswar, N., et al., *Comparative Protein Structure Modeling Using MODELLER*, in *Current Protocols in Protein Science*. 2001, John Wiley & Sons, Inc.

50. Wiederstein, M. and M.J. Sippl, *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins*, 2007. p. W407-W410.

51. Sippl, M., *Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures.* J Comput Aided Mol Des, 1993. **7(4)**: p. 473-601.

52. Sippl, M.J., *Calculation of conformational ensembles from potentials of mena force: An approach to the knowledge-based prediction of local structures in globular proteins.* Journal of Molecular Biology, 1990. **213**(4): p. 859-883.

53. Sippl, M.J., *Knowledge-based potentials for proteins.* Current Opinion in Structural Biology, 1995. **5**(2): p. 229-235.

54. Betz, S.F., et al., *Determination of the Binding Mode of Thienopyrimidinedione Antagonists to the Human Gonadotropin Releasing Hormone Receptor Using Structure-Activity Relationships, Site-Directed Mutagenesis, and Homology Modeling.* Journal of Medicinal Chemistry, 2006. **49**(21): p. 6170-6176.

55. Söderhäll, J.A., et al., *Antagonist and agonist binding models of the human gonadotropin-releasing hormone receptor.* Biochemical and Biophysical Research Communications, 2005. **333**(2): p. 568-582.

56. Hovelmann, S., et al., *Impact of Aromatic Residues within Transmembrane Helix 6 of the Human Gonadotropin-Releasing Hormone Receptor upon Agonist and Antagonist Binding* Biochemistry, 2002. **41**(4): p. 1129-1136.

57. Cook, J.V.F. and K.A. Eidne, *An Intramolecular Disulfide Bond between Conserved Extracellular Cysteines in the Gonadotropin-Releasing Hormone Receptor Is Essential for Binding and Activation*, 1997. p. 2800-2806.

58. De Lean, A., J.M. Stadel, and R.J. Lefkowitz, *A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled beta-adrenergic receptor.* Journal of Biological Chemistry, 1980. **255**(15): p. 7108-7117.

59. Samama, P., et al., *A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model.* Journal of Biological Chemistry, 1993. **268**(7): p. 4625-4636.

60. Rasmussen, S.G.F., et al., *Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex.* Nature, 2011. **advance online publication**.

Chapter 4. Theoretical and experimental studies on GnRH and agonist and antagonist peptide structures.

4.1 Introduction

The GnRH receptor is a clinical target for a number of therapies including prostate cancer and precocious puberty[1, 2]. However since there are currently no approved non-peptide drugs which are used to treat these conditions there is a need for rapid screening techniques in order to find possible small molecule GnRH-R pharmacophores. As the X-ray structure for GnRH-R is not currently available homology modelling must be used in combination with mutagenesis data to create a dockable receptor structure.

In this study the models were created using two known GPCR crystal structures; the β_2-adrenergic receptor and rhodopsin. When these models were created the high resolution β_2-adrenergic receptor was the only ligand bound GPCR crystal structure[3]. As bovine rhodopsin has a constitutively bound ligand[4], ligands known to bind GnRH-R were compared between both models and the better docking structure used for futher in silico experiments.

The traditional line between agonist and antagonist is slightly blurred when considering GPCRs. Traditionally an agonist initiates a specific response whilst an antagonist blocks this response, however continuous stimulation of GnRH-R by the GnRH-R peptide and the super-agonists causes desensitisation and internalisation of the receptor which causes cessation in signalling activity[5]. As noted in Chapter 1 the C-terminal tail which controls the rapid internalisation of the type II GnRH receptor

is missing from the type I structure. This causes a slower desensitisation response which is a poorer strategy to inhibit GnRH-R signlling; not just because of the increased time period to reach full efficacy but also because the initial surge or "flare effect"[6] in stimulation is detrimental to hormone dependent cancer treatment. Thus the GnRH-R antagonists were developed through residue substitution from the GnRH I peptide which had very high affinity for the receptor and demonstrated far quicker hypogonadism with no flare effect.

4.2 Comparisons of GnRH receptor agonist and antagonist peptides

In order to test the accuracy of the models an initial screen was performed which compared the docking response of four clinically approved antagonist peptides (Figure 4-1 and 4-2, Table 4-1). It was expected that these would score fairly well in docking simulations assuming our model was accurate. However a major difficulty in such an experiment is that small peptide structures are highly flexible and difficult to model which presents a challenge when carrying out docking work[7].

Table 4-1. Properties of the four selected high potency GnRH-R antagonist peptides for docking

Peptide name	Affinity (nM)	Molecular Weight	Clinical application	Hydrogen bond donors	Hydrogen bond acceptors	Reference
Abarelix	0.1	1416	Treatment of advanced prostate	13	16	[8]

			cancer			
Antide	<1	1591	Fertility treatment	14	18	[9]
Ganirelix	0.4	1570	Fertility treatment	16	20	[10]
Cetrorelix	0.2	1431	Treatment of hormone dependent cancers	17	18	[11]

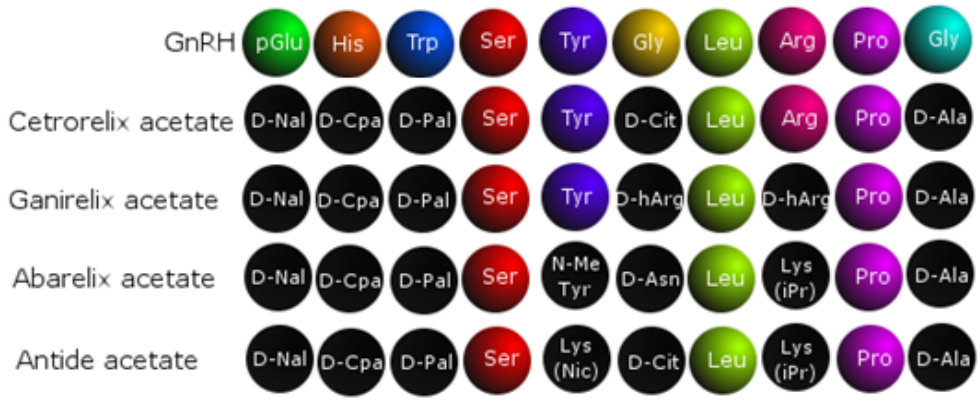


Figure 4-1. Comparison of 4 antagonist peptide structures with GnRH-I shows that a consensus of 3 key amino acids are required for binding to the receptor whilst avoiding activation of signalling. Non-natural amino acid substitutions are highlighted in black, terminology is as follows; D-Nal = D-Naptylalanine, D-Cpa = 4-chloro-phenylalanine, D-Pal = D-3-(3'-pyridyl)-alanine, D-Cit = D-Citruline, D-Ala = D-Alanine, D-hArg = dialkyl-D-homoarginine, N-Me Tyr = N-Methyl Tyrosine, D-Asn = D-Asparagine, Lys (iPr) = Lysine (N-Isopropyl), Lys (Nic) = Lysine (Nicotinoyl).

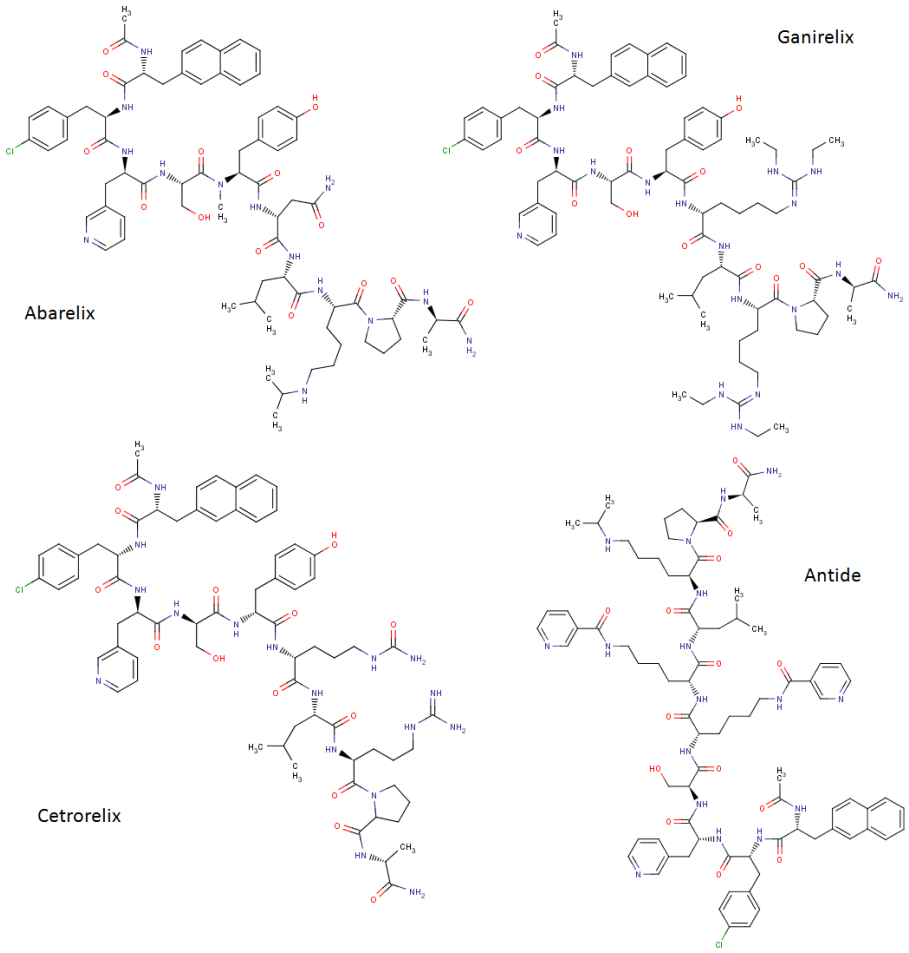


Figure 4-2. 2D structures of each antagonistic peptide. The structures show high similarity between all four peptide antagonists despite numerous residue substitutions.

The four antagonist peptides all share the Ser^{4}, Leu^{7} and Pro^{9} positions. However further substitution for non-natural amino acid groups is a key component of increasing binding affinity over the GnRH peptide. In contrast the agonist peptides that we examined (Table 4-2) have a much higher similarity to the GnRH-I sequence with minimal substitution in any of the four peptides.

The most successful GnRH agonist and antagonist peptides were discovered from substituting the GnRH-I sequence with non-natural amino acids. The effect of a

single substitution is highly variable and may cause a change in interaction between the residue and the pocket or by altering the shape of the peptide in a specific way[12]. Indeed it has been shown that from the entire ten residues of the decapeptide only a few key residues are required to preserve agonist activity (Figure 4-3 and 4-4).

Table 4-2. Properties of the four selected high potency GnRH-R agonist peptides for docking

Peptide Name	Affinity (nM)	Molecular weight	Clinical application	Hydrogen bond donors	Hydrogen bond acceptors	Reference
Goserelin		1329.4	Treatment of breast cancer	18	18	[13]
Histerelin	<1	1323.5	Treatment of precocious puberty	16	17	[14]
Leuprolide	<1	1269.4	Treatment of prostate cancer	16	16	[15]
Nafarelin	<0.1	1322.4	Treatment of endometriosis	17	17	[16]

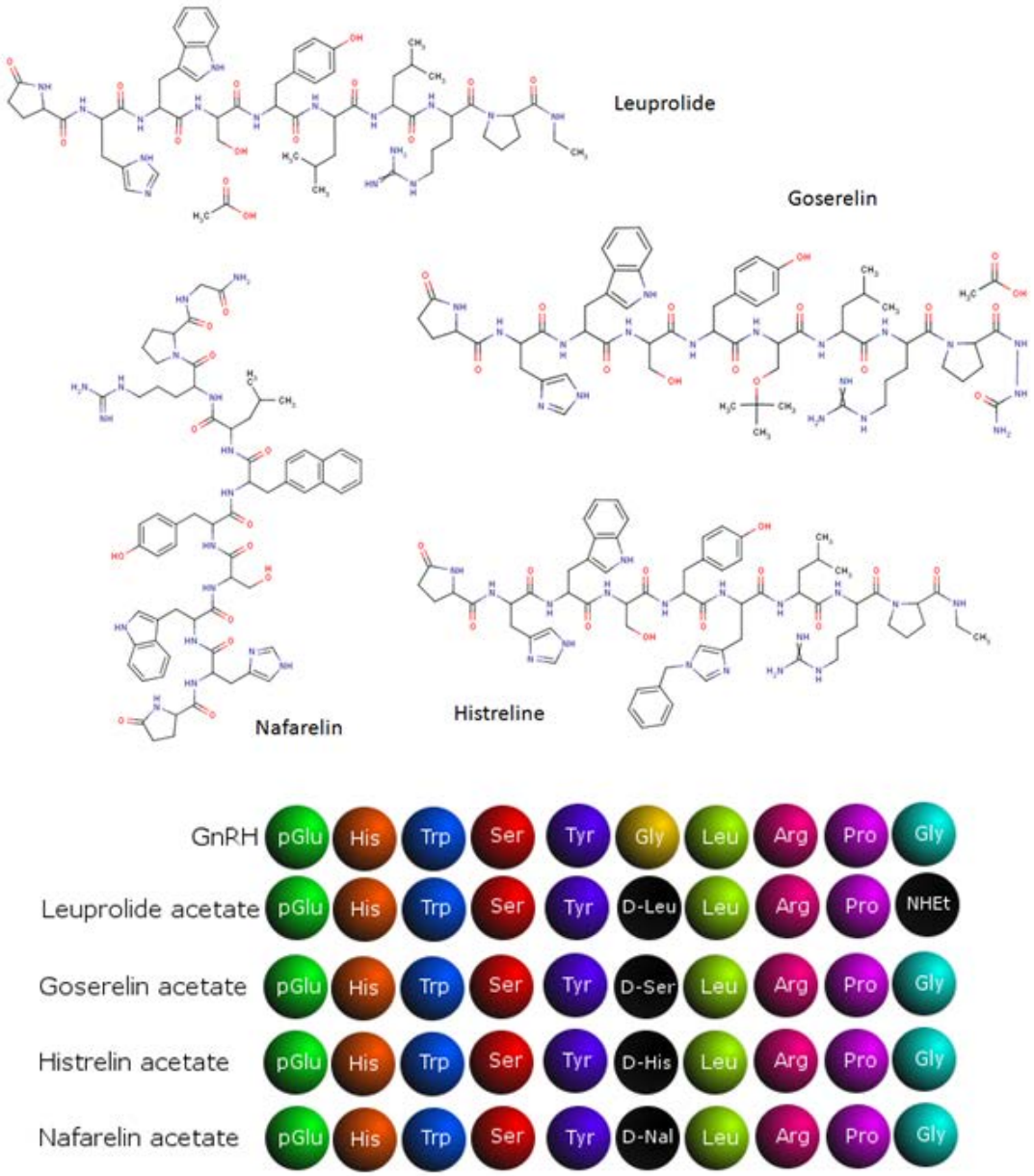


Figure 4-3. Structural comparison of four high potency GnRH analogue agonist peptides. Minimal substituitions separate the agonist peptides.

The almost exclusive substitution of residue Gly^6 is a means to further "lock" the peptide in a hairpin turn configuration which makes it more suitable for receptor activation. This method of peptidometrics was so successful that it gave rise to the so called GnRH "Superagonists" with binding affinities many hundreds of times stronger than the native GnRH-I[17],

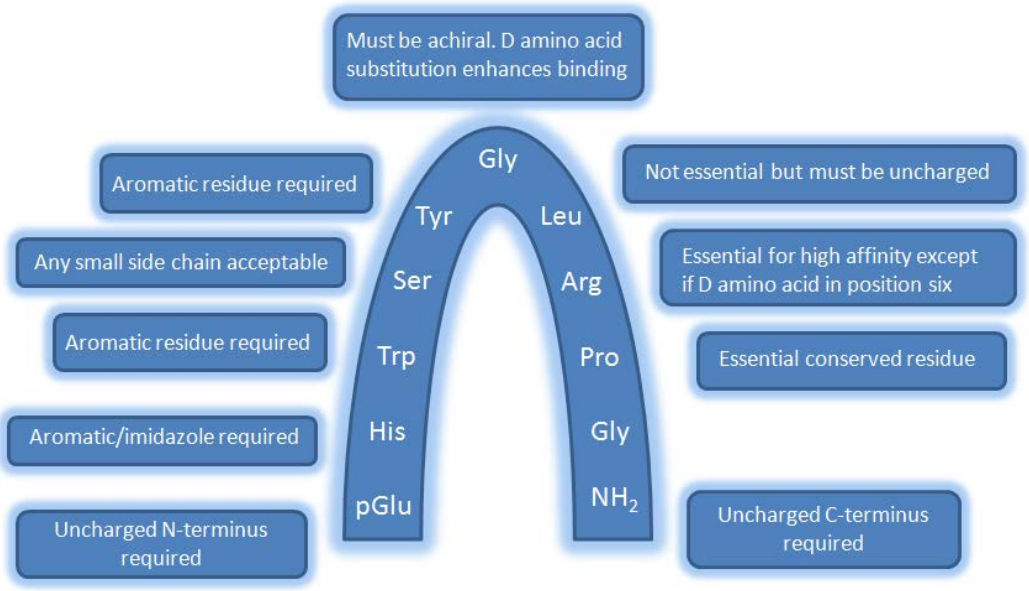


Figure 4-4. GnRH-I residue map and how residue substitution affects binding. Mutation of residues has shown that much of the decapeptide can be altered whilst still retaining functionality providing certain rules are followed to preserve specific interactions. Figure modified from [18].

4.3 Generating multi-conformers of peptide antagonist/agonists.

4.3.1 Generation of Antagonist multi-conformer peptide structures

As the docking procedure will incorporate a rigid body docking method, a series of 50 energy minimised conformers were generated for each of the antagonist peptide structures to give more flexibility to the poses. The conformers were created from the energy minimised 2D SDF file of each peptide as found in the PubChem database[19] then converted to 3D and run through the conformer generating algorithm in the MarvinSketch 5.1 software (Figure 4-5)[20].

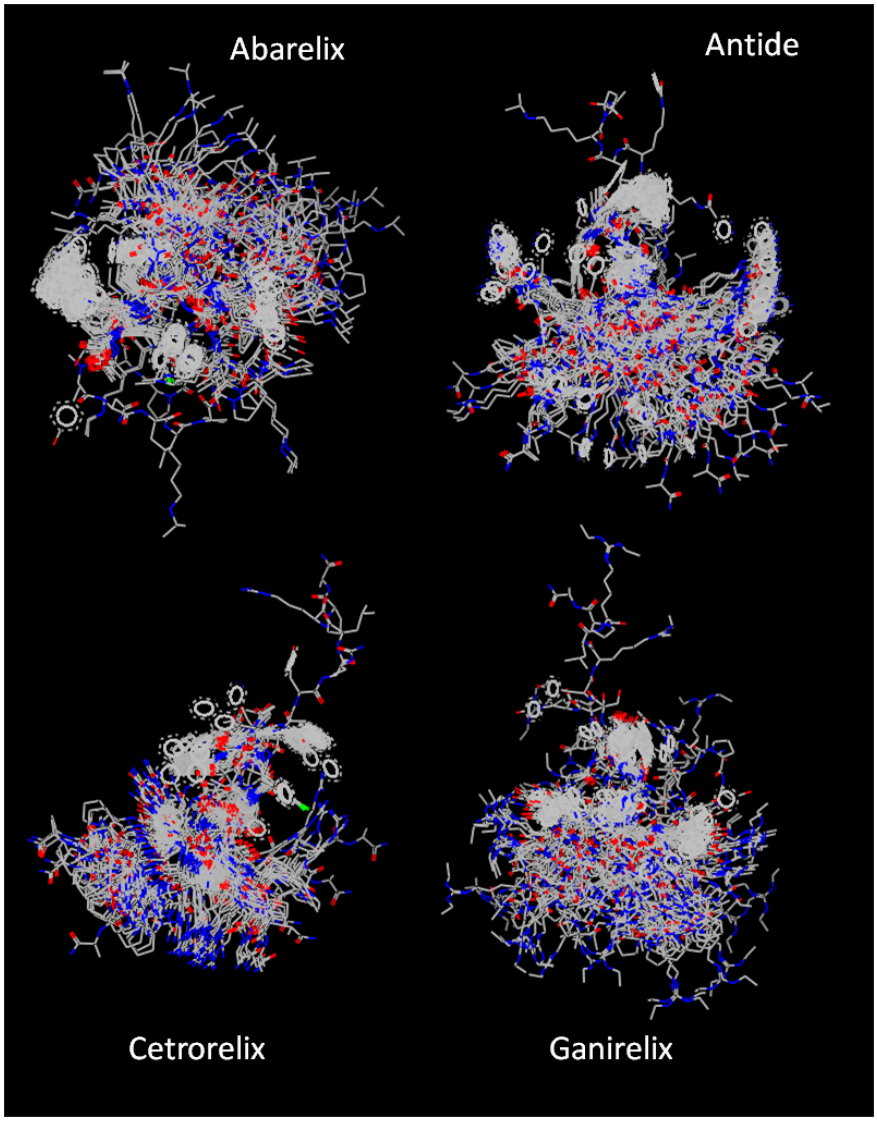


Figure 4-5. Superimposed conformer structures for each of the antagonist peptide. Abarelix Pubchem SID #17396902, Antide Pubchem SID #53788136, Cetrorelix Pubchem SID #11528755, Ganirelix Pubchem SID#50064278

4.3.2 Generation of agonist multi-conformer peptide structures.

The 2D structures for the agonist molecules were obtained from the PubChem database and the 3D structure and conformers (Figure 4-6) were created using MarvinSketch 5.1

4.4 Comparing the agonist and antagonist multi-conformer peptide structures

It can be noted that there seems to be a trend whereby the antagonist peptides form a hairpin which relaxes as further substitutions are made (Figure 4-7 dashed lines). However the agonist peptides seem far less prone to adopting this hairpin. This is particularly interesting given the structure of the GnRH peptide is believed to form the kind of hairpin bend shown by the antagonists from NMR data[21].

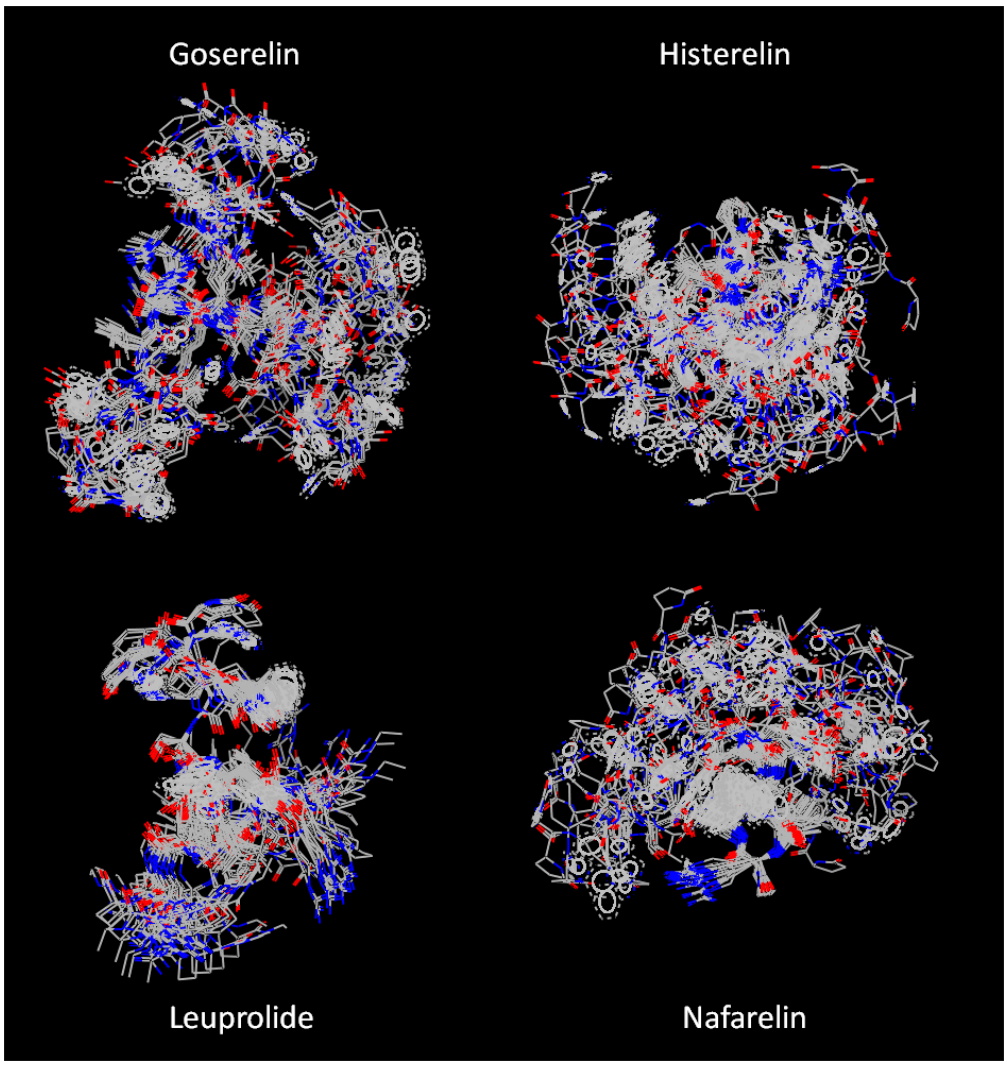


Figure 4-6. Superimposed conformer structures for each of the agonist peptides. Goserelin Pubchem SID #49989249, Histrelin Pubchem SID #49989266, Leuprolide Pubchem SID #577770, Nafarelin Pubchem SID #49989253.

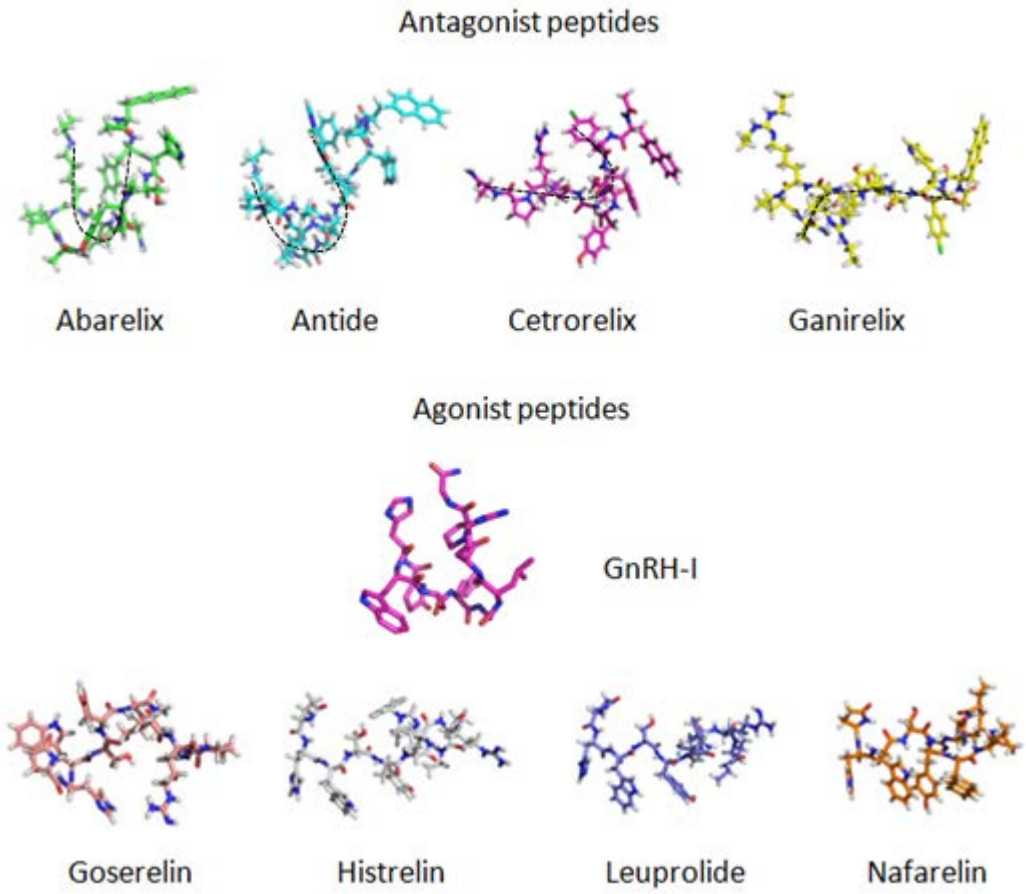


Figure 4-7. Comparison of energy minimised conformation of four antagonist and four agonist peptides against GnRH. There appears to be a trend which shows that agonist peptides are naturally more linear than the superagonists.

4.5 Docking agonist and antagonist peptide conformers using LIDAEUS and AutoDock vina

After generating a model structure it is useful to conduct a positive control by docking ligands into the active site which have already been established to bind by biophysical techniques.

In this section the docking of four commercial agonist and four commercial antagonist peptides into the rhodopsin and β_2 adrenergic receptor models are discussed using rigid body and flexible docking programs. Furthermore, as a

negative control, beta blocker structures are docked into both receptor models and then assayed for activity. This is because beta blockers are a β_2 Adrenergic receptor antagonist.

4.5.1 LIDAEUS and AutoDock vina allow for rigid body and flexible docking environments

LIDAEUS is a rigid body docking program which runs as a series of modules as described in Chapter 2.4. The modules create a docking and scoring pipeline which reads in a structure, docks it into the target protein and generates a score.

The process requires generating a series of site points which tell LIDAEUS where the atoms reside. This involves specifying which atoms are involved in interactions within the active site and also creating energy maps which describe the hydrophobic, hydrogen bond acceptor, hydrogen bond donor and buriedness characteristics of the site points.

AutoDock vina has a similar set up protocol (see Chapter 2.5 for full details), however rather than creating site points and energy maps a spacial grid is created which can accommodate the active site and ligand molecule. AutoDock vina has the advantage of being able to calculate potentials using flexible ligand and protein residues simultaneously. The force fields that it generates are similar to those in LIDAEUS, however it takes account of dispersion/repulsion, hydrogen bonding, electrostatics and desolvation.

4.5.2 Preparing the model for docking

To prepare the protein structure for molecular docking a modification was necessary. As previously observed the tertiary structure of the extracellular loop regions of the

GnRH-R model are likely to be highly variable and where they will be positioned cannot be determined with confidence. It would be undesirable for LIDAEUS to predict ligand interactions with the loop regions and so they were removed from the structure (Figure 4-8).

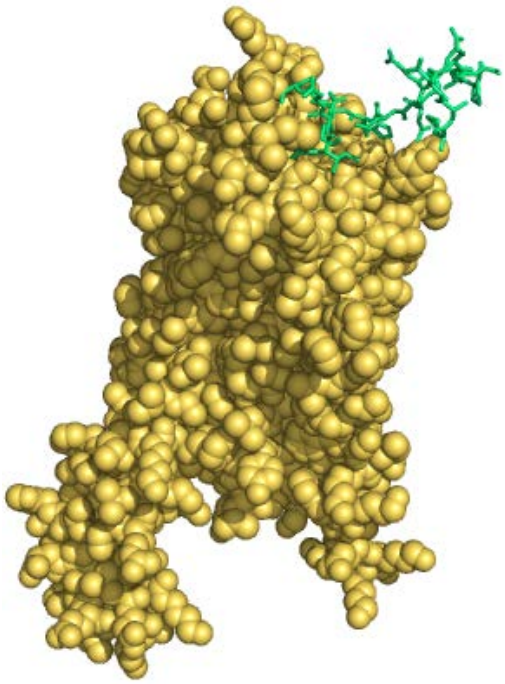


Figure 4-8. GnRH-R model structure used for docking. The residues represented by sticks were removed from the final docking model to reduce the possibility of incorrect results as a consequence of the flexibility of the region.

In addition, given that the key residues are all located on the helical regions the study would focus primarily on these. There are seven residues that are described as having a role in the binding of GnRH to its receptor from mutagenesis studies[22]: Asp98, Asn102, Asp302, Trp101, Lys121, Asn212 and Tyr290 therefore the docking experiment focussed around these regions.

4.5.3 Ligand site points specify binding regions on the receptor model

Using the GnRH-R model with extracellular loop regions removed, a set of site-points for LIDAEUS docking were created. Inspection of the binding pocket and knowledge of the key residues indicates that the pocket is relatively hydrophilic. However the Tyr290 and Trp101 residues will also confer some hydrophobic interactions. It was necessary to experiment with the site-point creation, as different parameters will produce a different balance of site-points types. It is desirable to create a set of site-points that reflect the nature of the binding pocket. The site-points are positioned around the binding pocket, close to the key residues as we might expect docked poses to form interactions with these residues (Figure 4-9).

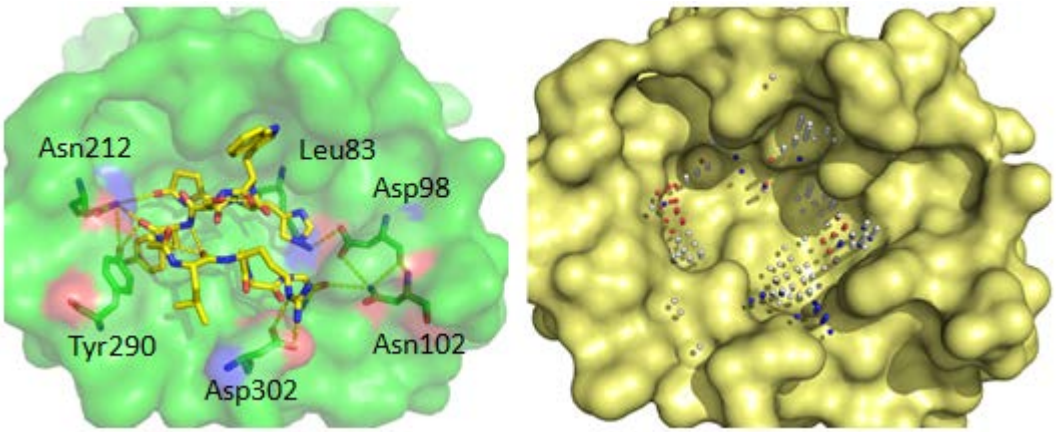


Figure 4-9. Site-points created for the GnRH receptor. The GnRH peptide is shown as yellow sticks. The site-points are displayed on the right. They are based on the physiochemical nature and structure of the binding pocket and use the GnRH peptide as a template to dictate where the site-points should be positioned.

4.5.4 Agonist and antagonist docking results using flexible poses in AutoDock vina using the β_2 adrenergic receptor homology model

Docking results for the agonist peptides provided a maximum binding energy of -8.8 Kcal/ mol for the Goserelin and Histrelin peptides. The range of affinities was -7.7 to -8.8 Kcal/ mol (Figure 4.10) which fits quite well with the experimentally observed affinity range for the antagonist peptides. The affinity for the antagonists (Figure 4-11) were slightly lower at -6.8 to -7.9 Kcal/ mol. This is slightly less than the agonist docking given that they have similar K_d values (Tables 4-1 and 4-2).

Peptide	Binding energy (Kcal/ mol)
Goserelin	-8.8
Histrelin	-8.8
Leuprolide	-8.2
Nafarelin	-7.7

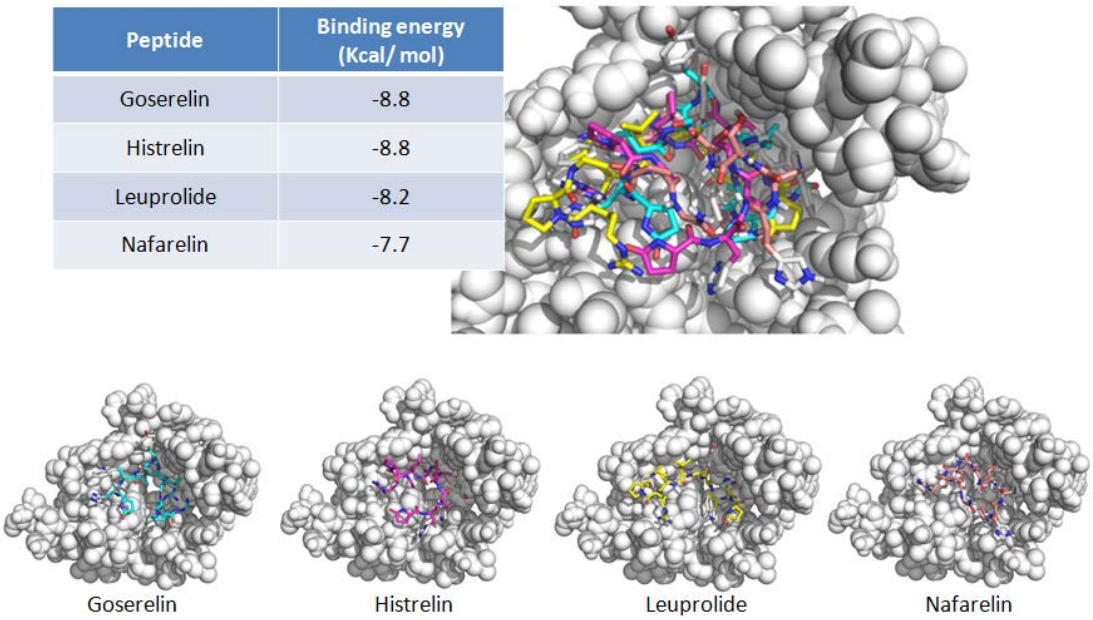


Figure 4-10. Overlay of four agonist peptides in their respective docked conformations using AutoDock vina. All adopt a similar spacial orientation which involves the sixth residue of the modified peptide occupying the twelve o'clock position. Peptides shown are Goserelin(Cyan), Histerelin(Magenta), Leuprolide(Yellow) and Nafarelin(Pink).

Peptide	Binding energy (Kcal/ mol)
Abarelix	-7.9
Antide	-7.9
Cetrorelix	-7.1
Ganirelix	-6.8

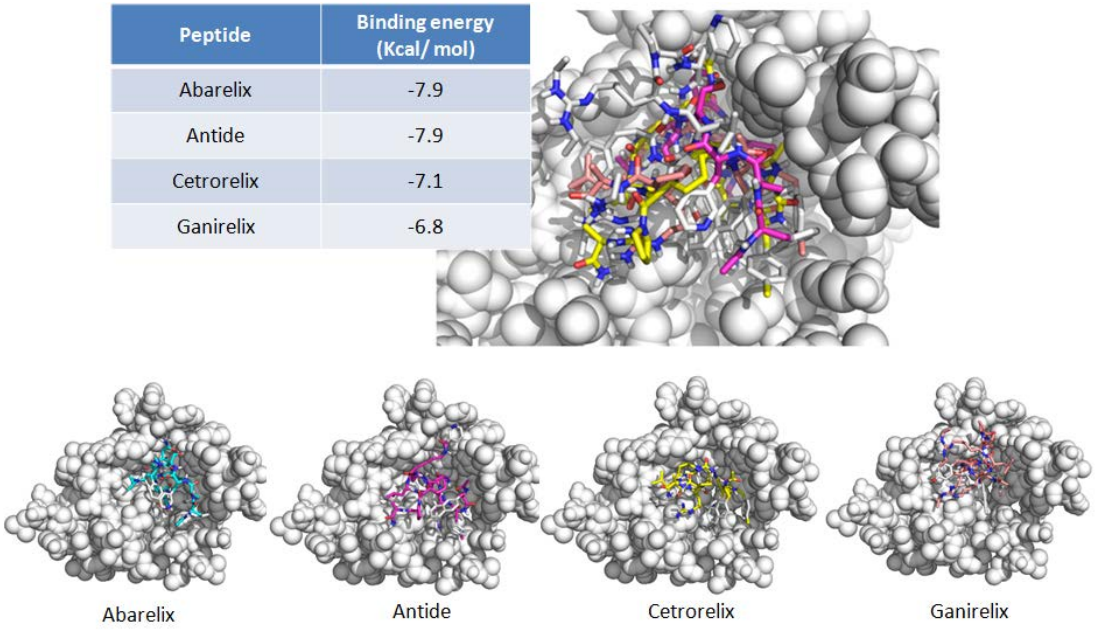


Figure 4-11. Overlay of four antagonist peptides in their respective docked conformations using AutoDock vina. The peptides all adopt a similar special orientation which involves the sixth residue of the modified peptide occupying the twelve o'clock position. Peptides shown are Abarelix(Cyan), Antide(Magenta), Cetrorelix(Yellow) and Ganirelix(Pink).

Examination of the hydrogen bonding interactions between the agonist peptides (Figure 4-10) indicates that there are numerous interactions for each docked peptide. Analysis of the residues within the GnRH receptor shows two residues in particular which are common features of the 4 peptide docking poses. Three of the four peptides interact with Tyr211 and all four interact with Asn305. Asn212 and Arg38 also have two peptides in common.

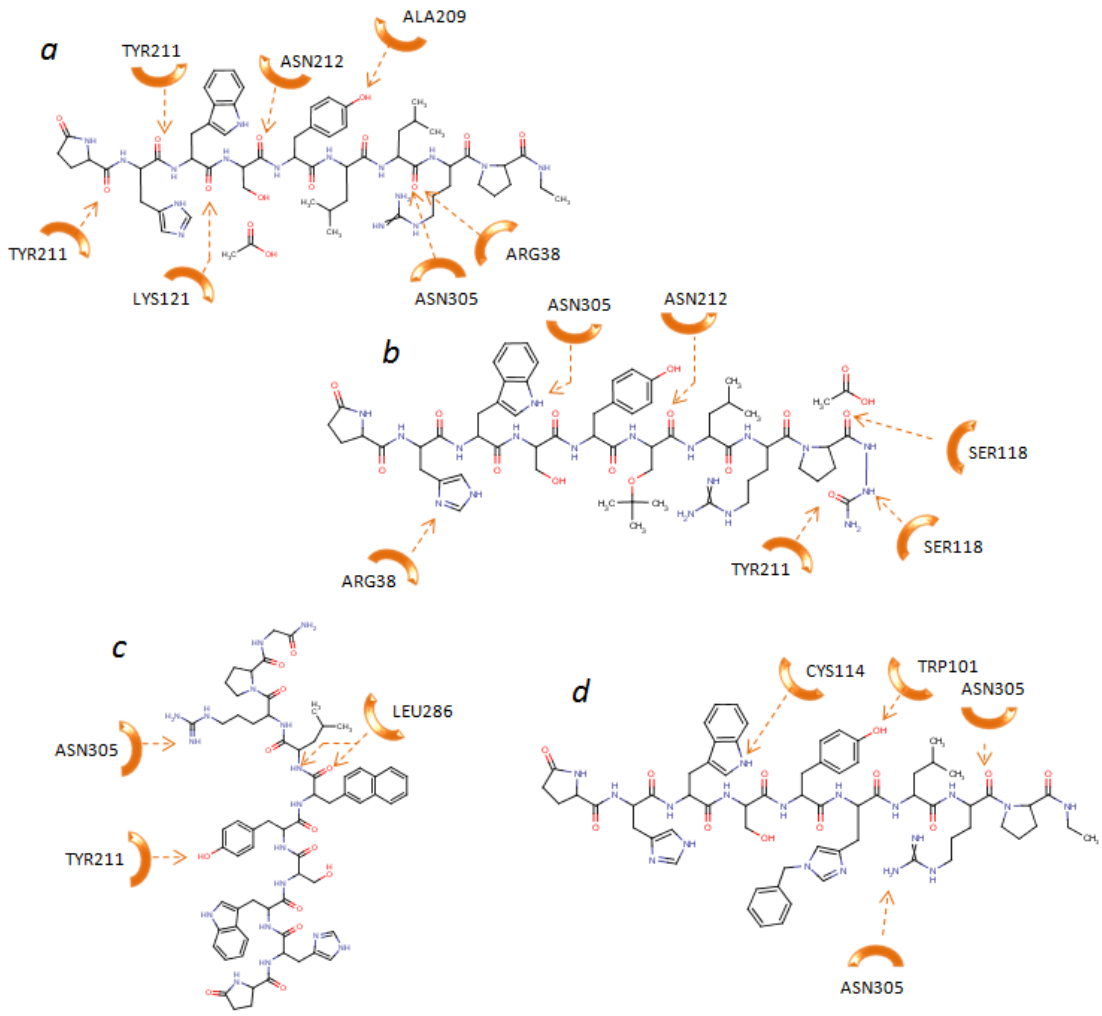


Figure 4-12. Potential interactions between docked agonist peptides and the gonadotrophin releasing hormone receptor using a β_2 adrenergic receptor derived model. (*a*) Leuprolide (*b*) Goserelin (*c*) Nafarelin (*d*) Histreline.

If the antagonist docking interactions are examined it can be seen that in this case there is no one residue that all four peptides have a common interaction with. Instead three peptides interact with Asn212, thereafter interacting residues share between one and two peptides in common. The residues involved in the docking studies have been shown in Figure 4-14 to allow visualisation of the hotspots in the receptor template.

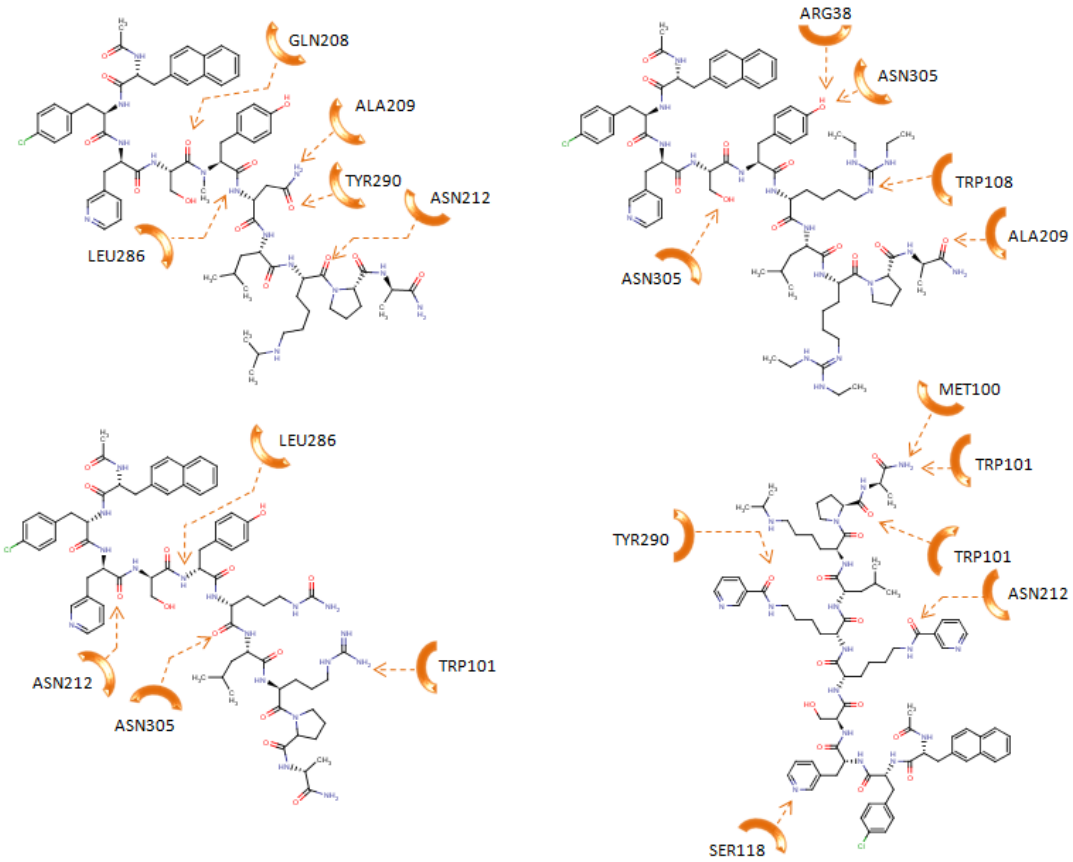


Figure 4-13. Potential interactions between docked antagonist peptides and the gonadotrophin releasing hormone receptor using a β_2 adrenergic receptor derived model. (*a*) Abarelix (*b*) Ganirelix (*c*) Ceterorelix (*d*) Antide.

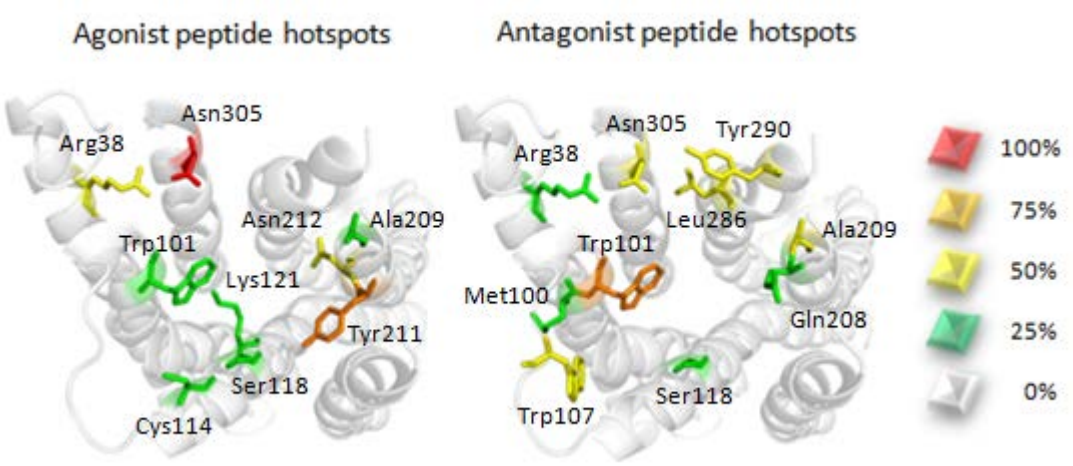


Figure 4-14. Identifying hotspots in agonist and antagonist peptide binding to the GnRH-R β_2 adrenergic receptor homology model. The colour scale indicates the percentage of docked peptides which interact with the highlighted residue.

Comparison of the hotspots between agonists and antagonists highlights five residues which appear in both templates, Arg38, Trp101, Asn212, Ala209 and Asn305. These were compared to the GnRH binding residues which have been identified by site directed mutagenesis (Figure 4-9) and were found to have Trp101 and Asn212 in common. Whilst Asn305 has not been shown by mutagenesis to be involved in GnRH binding, there are no reports in the literature where it was tested. As Asn305 is close to Asp302 which is part of the GnRH binding site, there is a possibility that it could perform an as yet undiscovered function.

Therefore from the docking results using this model it could be advantageous to select small molecules which are able to bind to these residues in the development of future therapeutics.

4.5.5 Agonist and antagonist docking results using flexible poses in AutoDock vina using the rhodopsin homology model

The validation results from the β_2 adrenergic receptor homology model showed good scoring for the clinically approved peptides. Therefore the same validation method was applied to the GnRH-R rhodopsin homology model for both the agonist (Figure 4-15) and antagonist (Figure 4-16) peptides.

Peptide	Binding energy (Kcal/ mol)
Goserelin	-7.5
Histrelin	-7.3
Leuprolide	-8.0
Nafarelin	-8.6

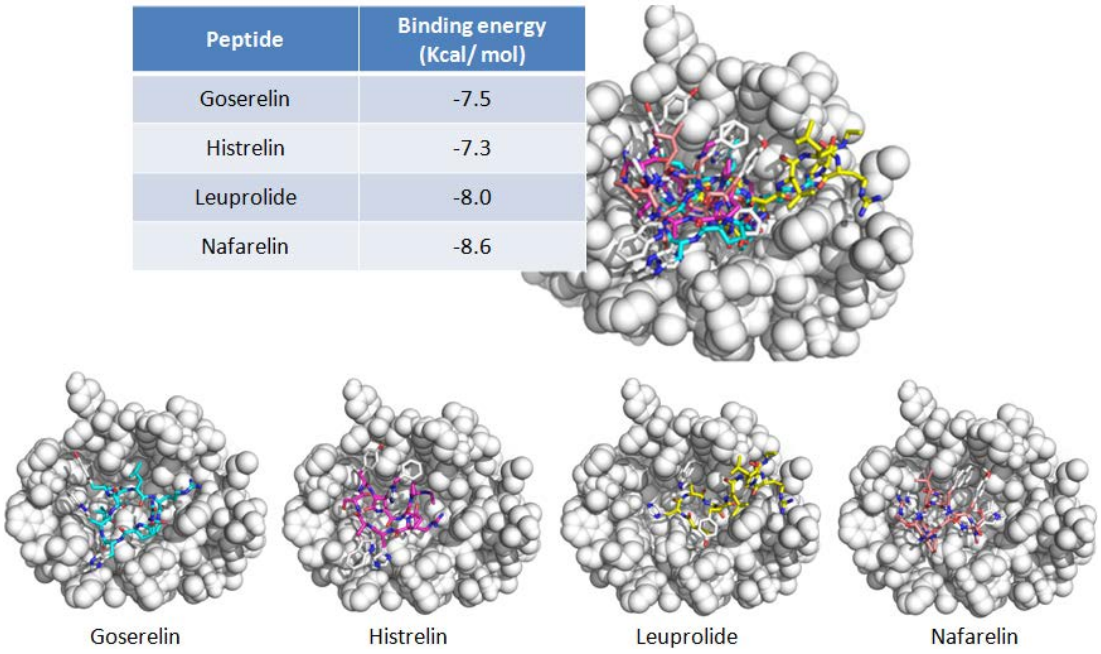


Figure 4-15. Overlay of four agonist peptides in their respective docked conformations using AutoDock vina. All adopt a similar spacial orientation which involves the sixth residue of the modified peptide occupying the twelve o'clock position. Peptides shown are Goserelin(Cyan), Histerelin(Magenta), Leuprolide(Yellow) and Nafarelin(Pink).

Peptide	Binding energy (Kcal/ mol)
Abarelix	-7.2
Antide	-7.8
Cetrorelix	-8.0
Ganirelix	-7.2

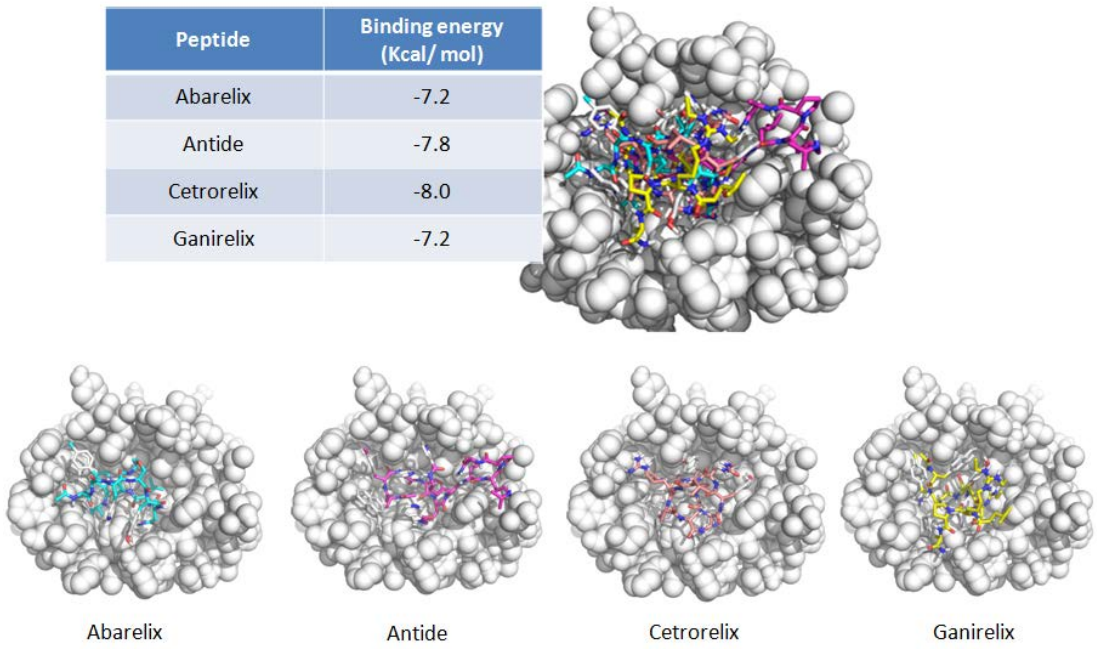


Figure 4-16. Overlay of four antagonist peptides in their respective docked conformations using AutoDock vina. The peptides all adopt a similar special orientation which involves the sixth residue of the modified peptide occupying the twelve o'clock position. Peptides shown are Abarelix(Cyan), Antide(Magenta), Cetrorelix(Yellow) and Ganirelix(Pink).

In this case the binding affinity range for the agonists was -7.3 to -8.6 Kcal/ mol whilst the range for the antagonists was similar at -7.2 to -8.0 Kcal/ mol. Whilst the increased consistency between both sets of results is more representative of the measured affinity of the peptides we should still consider that it is not necessarily an indication that the active site in the rhodopsin model is more or less accurate than that from the β_2 adrenergic receptor model test. Conclusions on model accuracy can only be made from *in vivo* assay results.

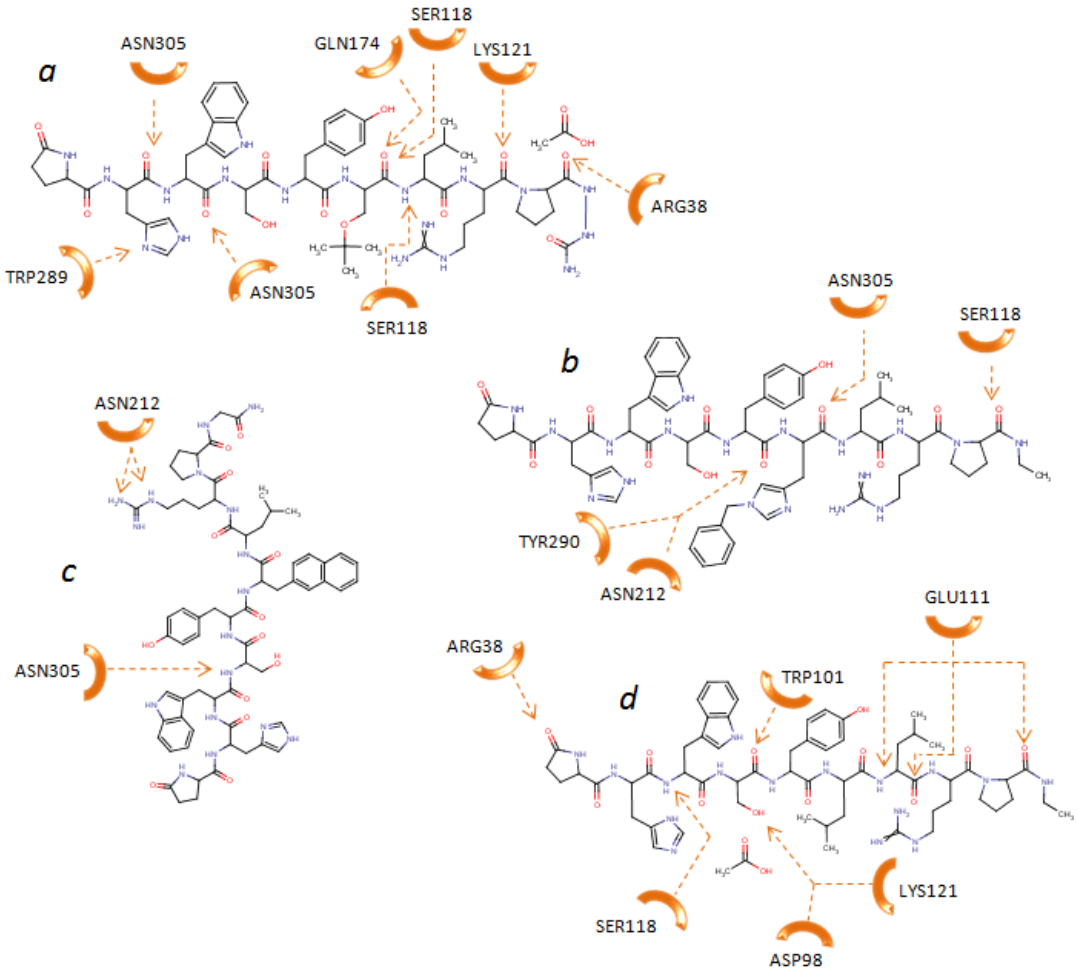


Figure 4-17. Potential interactions between docked agonist peptides and the gonadotrophin releasing hormone receptor using a rhodopsin derived model. (*a*) Leuprolide (*b*) Goserelin (*c*) Nafarelin (*d*) Histreline.

The possible interaction sites from the GnRH receptor to the modified peptides have been detailed in Figures 4-17 and 4-18. The agonist docking showed that both Asn305 and Ser118 were able to associate with three of the four peptides. If we compare that to the agonist results from the β_2 adrenergic receptor model we note that Asn305 features in both sets of results.

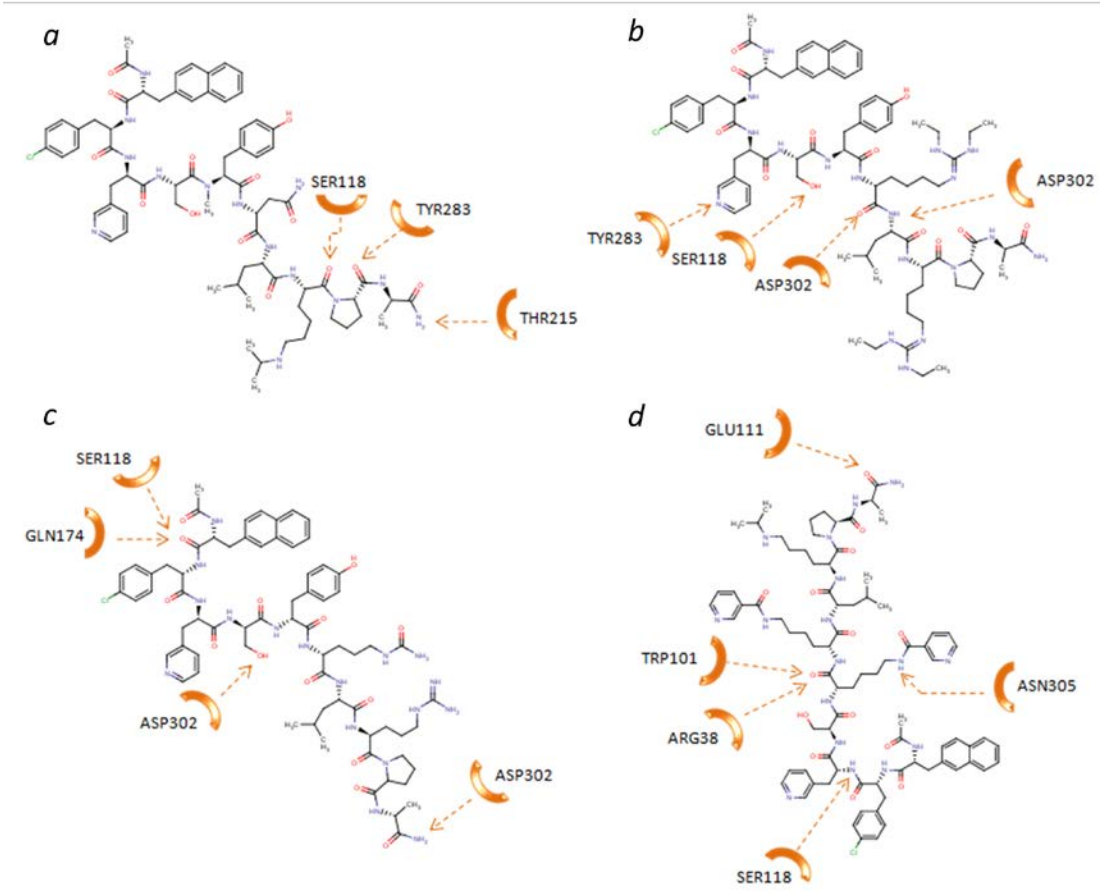


Figure 4-18. Potential interactions between docked antagonist peptides and the gonadotrophin releasing hormone receptor using a rhodopsin derived model. (*a*) Abarelix (*b*) Ganirelix (*c*) Ceterorelix (*d*) Antide.

The antagonist peptides heavily favoured interaction with Ser118 in the receptor active site, however this residue was not a significant contributor in the β_2 adrenergic receptor model. The hotspots of peptide binding are illustrated in Figure 4-19.

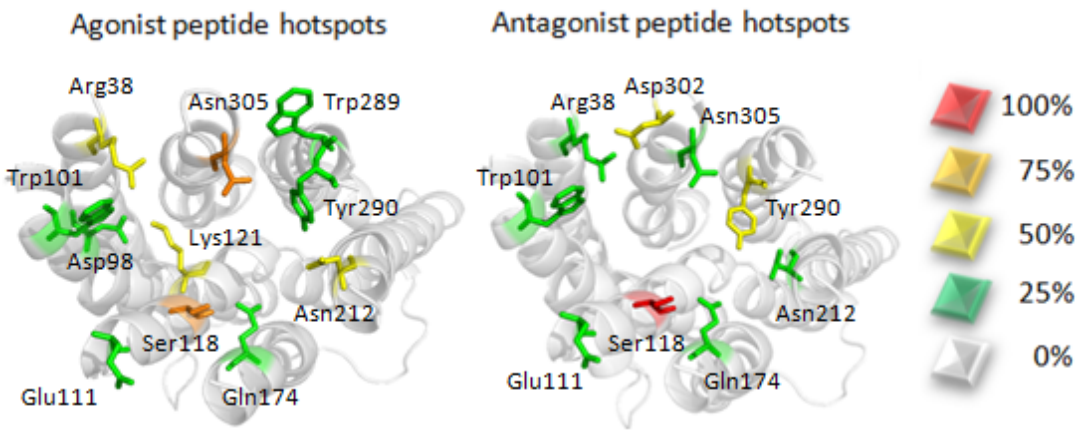


Figure 4-19. Identifying hotspots in agonist and antagonist peptide binding to the GnRH-R rhodopsin homology model. The colour scale indicates the percentage of docked peptides which interact with the highlighted residue.

A comparison between the docking modes of both peptide types shows 8 residues which appear in both cases, Arg38, Trp101, Glu111, Ser118, Gln174, Asn212, Tyr290 and Asn305. Of these amino acids, Trp101 and Asn212 are residues which have been shown to be key components of the GnRH binding site from mutational study.

The docking results from the β_2 adrenergic receptor model and the rhodopsin model showed that the known peptide agonist and antagonists scored well in both models. Furthermore a number of residues were highlighted which participated in interactions between multiple peptides and were observed in both model structure docking experiments.

The use of flexible docking peptide ligands into the GnRH-R receptor allowed the validation of the model structures as well as identifying potential residues for targeting with small molecules.

For further comparison both the β_2 adrenergic receptor model and the rhodopsin model were docked with the antagonist peptide structures using the rigid body LIDAEUS software.

4.5.6 Comparing Antagonist peptide rigid body docking results from LIDAEUS using rhodopsin and β_2 adrenergic receptor homology models

4.5.6.1 Antagonistic peptide docking into the Rhodopsin derived GnRH-R model

Taking the same 50 conformers that were described previously for Abarelix, Antide, Cetrorelix and Ganirelix the docked solutions within the GnRH-R active site were found using LIDAEUS (Table 4-3).

The scores for the peptides were fairly similar between all four structures with the exception of Cetrorelix which scored far more favourably. However upon closer inspection of the docked poses it was evident that LIDAEUS had difficulty docking the peptide structures deeply into the binding site of the model. In all cases the best solution that LIDAEUS provided showed the peptides sitting further outside the pocket than was shown by AutoDock vina.

Table 4-3. Scoring 50 peptide conformers for the Abarelix, Antide, Cetrorelix and Ganirelix antagonists using the rhodopsin derived GnRH-R model, scores shown are the top 10 of those obtained.

Abarelix LIDAEUS Score	Antide LIDAEUS Score	Cetrorelix LIDAEUS Score	Ganirelix LIDAEUS Score
-17.8	-16.6	-26.9	-18.0
-17.5	-16.5	-19.1	-17.4
-16.4	-15.1	-18.9	-16.0
-16.0	-15.0	-18.6	-16.0
-13.3	-14.4	-16.5	-14.7
-13.1	-13.5	-16.2	-13.5
-12.9	-13.1	-16.2	-13.2
-12.8	-9.7	-15.5	-9.5

4.5.6.2 Antagonistic peptide docking – β_2 adrenergic receptor model

The same antagonist conformers were then docked using the β_2 adrenergic receptor model (Table 4-4). This time the peptide structures generally scored more favourably than in the rhodopsin model study, however there was still a notable difference in how deeply the peptide was buried in the pocket compared to the flexible docking approach. This suggests that in docking studies where highly flexible ligands are to be used there is an advantage in using an approach where the ligand and protein are allowed to be flexible.

Table 4-4. Scoring 50 peptide conformers for the Abarelix, Antide, Cetrorelix and Ganirelix antagonists using a structure modelled from the β_2 adrenergic receptor, scores shown are the top 10 of those obtained.

Abarelix LIDAEUS Score	Antide LIDAEUS Score	Cetrorelix LIDAEUS Score	Ganirelix LIDAEUS Score
-25.5	-18.8	-23.0	-27.1
21.2	-18.1	-22.2	-21.6
-20.3	-17.6	-19.8	-19.5
-15.7	-15.1	-19.7	-19.1
-13.4	-15.1	-19.6	-17.2
-13.1	-14.9	-19.3	-16.6
-12.6	-13.9	-18.0	-15.6
-12.3	-13.7	-16.5	-15.3
-12.2	-13.4	-16.4	-15.0
-12.0	-13.2	-15.5	-14.8

4.6 Crystallisation and characterisation of GnRH peptide

A number of studies have examined the structure of the GnRH peptides using a number of characterisation techniques such as circular dichroism, NMR, fluorescence and a number of computational modelling methods[23-34]. A β-turn at position Gly6 was a common feature of all studies, however due to the highly flexible nature of the peptide no single consensus structure has been solved, if a single conformation in solution even exists. Hence there is considerable interest in obtaining a crystal structure of the GnRH peptide.

4.6.1 GnRH I [D-Lys6,Trp7,Leu8] peptide crystallisation strategy

After examining the conditions in which a number of short peptides were crystallised (Table 4-5) a hanging drop method was decided to be the most successful approach.

Table 4-5. Crystallisation parameters for low molecular weight peptide fragments.

Peptide Sequence	Concentration (mg/ml)	Method used	Precipitant	Reference
NNQQNY	30	Hanging-Drop	100mM ZnSO4	[35]
GNNQQNY	10	Hanging-Drop	100mM NaCl	[35]
GNNQQNY	10	Bulk crystallization	N/A	[35]
NNQQ (form 1)	30-50	Hanging-Drop	20% w/v PEG 4000; 20% v/v 2-propanol)	[35]
NNQQ (form 2)	30-50	Hanging-Drop	35% v/v tert-butanol	[35]

VEALYL	1.4	shaking at 37 °C for 1 month	150 mM NaCl	[35]
LYQLEN	1.6	shaking at 37 °C for 1 month	150 mM NaCl	[35]
VQIVYK	30	Hanging-Drop	45% v/v MPD	[35]
MVGGVV (form 1)	30	Hanging-Drop	30% v/v MPD	[35]
MVGGVV (form 2)	30	Hanging-Drop	20% v/v 2-propanol	[35]
GGVVIA	15	Hanging-Drop	2.0 M ammonium sulfate	[35]
SSTSAA	30	Hanging-Drop	20% w/v PEG4000	[35]
SNQNNF	10	Hanging-Drop	200 mM sodium acetate	[35]
ELLKKLLEELKG	10	Hanging-Drop	80% ammonium sulphate	[36]
AAAK	1	shaking at 37 °C for 1 month	PBS	[37]
ECCNPACGRHYSC	45	Dialysis with PEG 20K	PEG 20000	[38]

The first hanging drop screen incorporated a simple ammonium sulphate screen to examine how easily the peptide could be precipitated from solution. A screen of 5-30% ammonium sulphate was used (Figure 4-20)

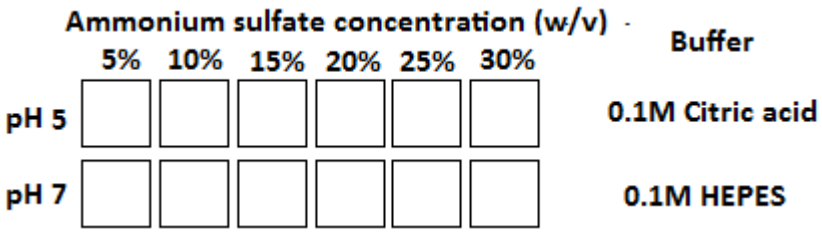


Figure 4-20. Preliminary ammonium sulphate screen to determine precipitation point.

From the ammonium sulphate screen it was found that at pH 5 the 5% drop remained clear however the other drops all showed a phased appearance which failed to develop into precipitate. In contrast at pH 7 the 5% and 10% drops both showed light precipitation whilst the 15%-30% drops showed heavy precipitation. A broad matrix screen was then used which consisted of the Crystal Screen 1 and 2 kits (Hampton Research) at temperatures of 4 °C and 17 °C using 10 mg/ ml peptide per drop.

From these 196 conditions there were 3 which showed signs of crystal growth (Table 4-6).

Table 4-6. Conditions which show microcrystal formation of the GnRH peptide.

Salt	Buffer	Precipitant	Temperature (°C)
0.2 M Zinc acetate dihydrate	0.1 M Sodium cacodylate trihydrate pH 6.5	18% w/v Polyethylene glycol 8,000	4
None	None	2M ammonium sulfate	4 and 17

None	0.1M sodium cacodylate trihydrate pH 6.5	1.4M sodium acetate trihydrate	17

These conditions were interesting in the respect that a screening condition which was identical to the zinc conditions but replaced zinc with calcium produced no microcrystal growth. This suggests that the zinc itself has a specific stabilising effect on the peptide which can lock the peptide into a particular conformation, perhaps through GnRH-I chelating with the metal. This compliments the observation that incubating GnRH with zinc prior to competition binding with busarelin reduces GnRH effectiveness[39].

Taking the conditions from Table 4-6 and creating matrixes around these conditions to improve crystal size was unsuccessful in producing larger crystals, however the initial hit condition containing 0.2 M Zinc acetate dihydrate yielded needle crystals (Figure 4-21) which were large enough to obtain a dataset of 54 images at Δphi of 7.5° (Table 4-7) using the IO3 beamline at the Diamond Light Source.

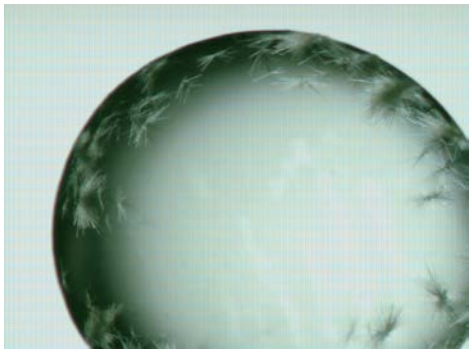


Figure 4-21. Needle clusters of GnRH I [D-Lys6,Trp7,Leu8] peptide in 0.2 M Zinc acetate dihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 18% w/v Polyethylene glycol 8,000 at 4 °C.

The data processing was performed using the program MOSFLM[40]. Owing to the weak nature of spots indexing was performed by manual selection of spots from multiple images throughout the dataset (Figure 4-22).

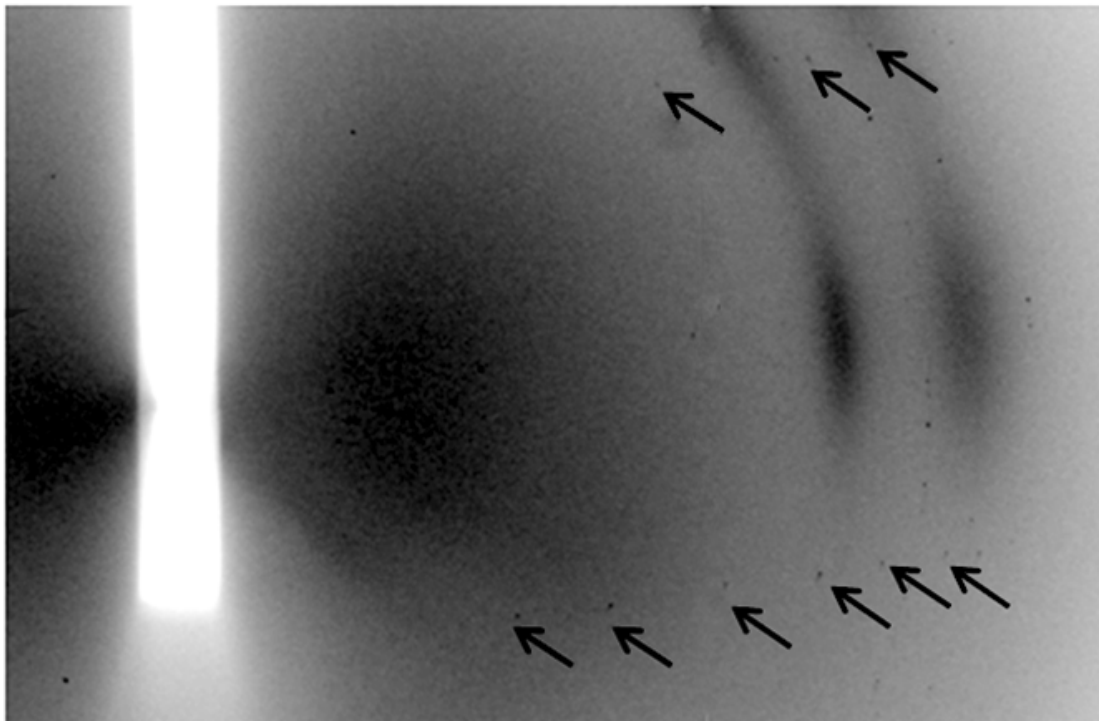


Figure 4-22. Diffraction image of GnRH I [D-Lys6,Trp7,Leu8] peptide needles. Black arrows highlight the spots which are difficult to observe.

Analysis using the program pointless[40] suggested the space group to be $P2_12_12_1$ and the data processing statistics are shown below in Table 4-7.

Table 4-7. Processing statistics for GnRH I [D-Lys6,Trp7,Leu8] peptide.

Space group	$P2_12_12_1$
R sym	18.7%
Resolution	3.0 Å
Completeness	98%
Cell dimension	A 11.5 α 90° B 39.8 β 90° C 43.3 γ 90°

Calculating the cell volume gives a value of 19818Å^3 and the molecular weight of the peptide is 1283Da. Using a Matthews coefficient[41] (Equation 1) this would give a solvent content as indicated below in Table 4-8.

$$Vm = \frac{Cell\ Volume\ (cubic\ angstrom)}{M*NA*NM} = \frac{V}{M*Z} \qquad \text{(Equation 1)}$$

M = molecular weight of protein in daltons

V = volume of unit cell.

Z = number of molecules in unit cell = NA*NM

NA = number of asymmetric units

NM = number of molecules in asymmetric unit.

Table 4-8. Calculating solvent content of the asymmetric unit.

Peptides in asymmetric unit	Matthews coefficient	% solvent
1	3.86	68.15
2	1.93	36.30

Searching the Cambridge Structural Database (CSD)[42] for small peptides yielded 20 structures with 12-17 residues. From this set of peptides, 2 structures (Figure 4-23 and Table 4-9) also had a space group of $P2_12_12_1$. Both of these structures adopted α-helical conformations however they both featured a number of α-aminoisobutyryl (Aib) residues which are a facilitator of helix formation[43, 44].

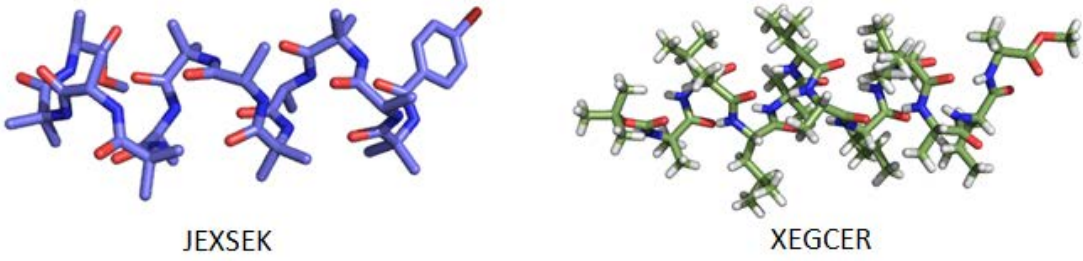


Figure 4-23. 2 peptide structures with space group $P2_12_12_1$ were found in the CSD. Both structures adopt α-helical conformations due to α-aminoisobutyryl inclusion within the peptide sequence.

Table 4-9. Cell dimension parameters for $P2_12_12_1$ peptides from the CSD.

	JEXSEK	XEGCER
Molecular weight	1155	1321
Space group	$P2_12_12_1$	$P2_12_12_1$
Residue count	12	13
a	10.33	9.96
b	18.13	20.12
c	35.0	39.31
α	90	90
β	90	90
γ	90	90
Peptides in asymmetric unit	1	1
Matthews coefficient	1.42	1.49
Solvent (%)	13.31	17.5

From analysis of other peptide crystals of similar size this would suggest that there are 2 peptide molecules in the asymmetric unit. In addition a closer look at the unit cell dimensions indicate that the peptides are less likely to be helical and more likely to be strands.

Attempts to solve the structure of the peptide using segments of β-strand have so far been unsuccessful. This is likely due to the weakly diffracting nature of the needle

crystals obtained. To date no better crystals have been obtained however further trials using the best condition found may yield better results in the future.

4.6.2 Characterisation of GnRH peptide by circular dichroism

Circular dichroism has been used before to establish structural information on the GnRH I peptide, however the GnRH I [D-Lys6,Trp7,Leu8] is thought to be a slightly more stable form of the peptide. Structural features were analysed using CD with increasing concentrations of 2,2,2-trifluoroethanol (TFE) (Figure 4-24).

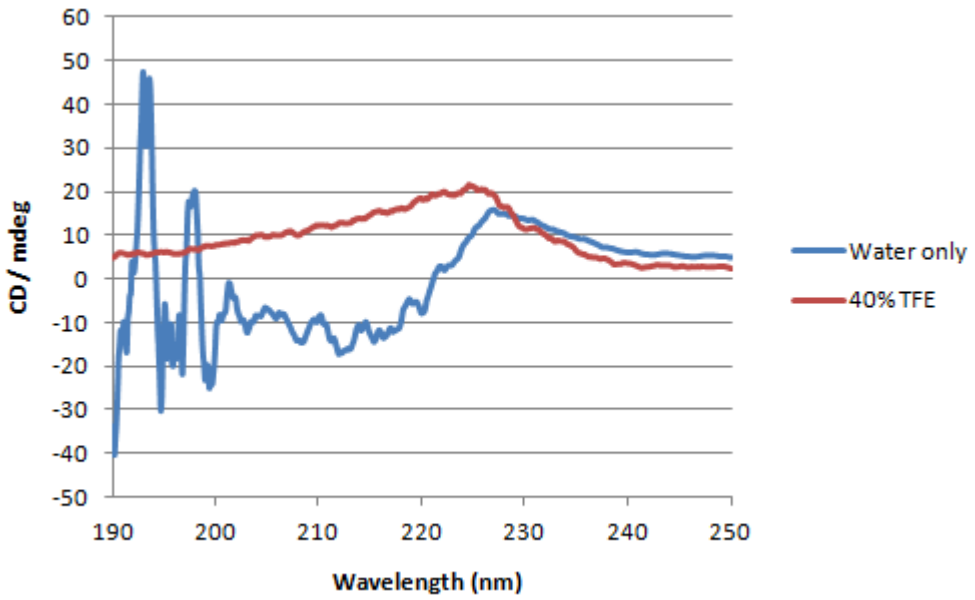


Figure 4-24. Circular dichroism spectra of GnRH I [D-Lys6,Trp7,Leu8] in water and TFE solution.

All spectra were recorded using a nitrogen flushed Jasco J-810 spectropolarimeter. In each case a final GnRH I [D-Lys6,Trp7,Leu8] concentration of 1mg/ ml was used which was prepared directly from lyophilised stock. Cell path length used was 0.02 cm and data was recorded from 190 to 250 nm in 0.1 nm steps and was accumulated over 3 runs which were corrected for buffer. The CD spectra was deconvoluted using the DICHROWEB server[45].

After analysis of the CD spectra it was found that in pure water the peptide was 87.5% unordered with minor regions of strand, partial helix and turn. After addition of 40% TFE the unordered regions were 24%, strand was 50% and turn 22%. Helix contribution was a little over 4%. The large unordered regions of the peptide in water are an expected result, environment plays a large part in determining secondary structure[46]. Short peptide sequences are more flexible in water where stabilising interactions with larger molecules are not available[47]. Adding TFE to peptide solutions has been shown to cause a stabilising effect[48], a contributing factor as to why this is observed is thought to be TFE weakening the interactions between amide bonds and water which increases intramolecular hydrogen bonding[49, 50].

The native structure of the GnRH I peptide is thought to form a natural hairpin in physiological conditions[51]. Studies have shown that formation of a hairpin is a stabilising effect which is associated with proteins with structures composed of high proportions of strands and turns[52-54]. The CD data obtained in the 40% TFE study shows strand and turns to be the most common secondary structure elements within the peptide which could suggest that GnRH I [D-Lys6,Trp7,Leu8] could also adopt this hairpin configuration under these conditions.

4.8 References

1. D. Dondi, C.F., M. Piccolella, M. Bologna, M. Motta, *GnRH agonists and antagonists decrease the metastatic progression of human prostate cancer cell lines by inhibiting the plasminogen activator system.* Oncology Reports, 2006. **15**(2): p. 393-400.

2. Sullivan, S.K., et al., *Kinetics of nonpeptide antagonist binding to the human gonadotropin-releasing hormone receptor: Implications for structure-activity relationships and insurmountable antagonism.* Biochemical Pharmacology, 2006. **72**(7): p. 838-849.

3. Rasmussen, S.G.F., et al., *Crystal structure of the human β2 adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-387.

4. Palczewski, K., et al., *Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor.* Science, 2000. **289**(5480): p. 739-745.

5. McArdle, C.A., et al., *Signalling, cycling and desensitisation of gonadotrophin-releasing hormone receptors.* Journal of Endocrinology, 2002. **173**(1): p. 1-11.

6. Coccia, M.E., et al., *GnRH antagonists.* European Journal of Obstetrics & Gynecology and Reproductive Biology, 2004. **115, Supplement**(0): p. S44-S56.

7. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications.* Nat Rev Drug Discov, 2004. **3**(11): p. 935-949.

8. Mongiat-Artus, P. and P. Teillac, *Abarelix: the first gonadotrophin-releasing hormone antagonist for the treatment of prostate cancer.* Expert Opinion on Pharmacotherapy, 2004. **5**(10): p. 2171-2179.

9. Kakar, S.S., *Inhibition of Growth and Proliferation of EcRG293 Cell Line Expressing High-Affinity Gonadotropin-releasing Hormone (GnRH) Receptor Under the Control of an Inducible Promoter by GnRH Agonist (d-Lys6)GnRH and Antagonist (Antide).* Cancer Research, 1998. **58**(20): p. 4558-4560.

10. Jiang, G., et al., *GnRH Antagonists: A New Generation of Long Acting Analogues Incorporating p-Ureido-phenylalanines at Positions 5 and 6.* Journal of Medicinal Chemistry, 2000. **44**(3): p. 453-467.

11. Reissmann, T., et al., *The LHRH antagonist Cetrorelix: a review.* Human Reproduction Update, 2000. **6**(4): p. 322-331.

12. Huirne, J.A.F. and C.B. Lambalk, *Gonadotropin-releasing-hormone-receptor antagonists.* The Lancet, 2001. **358**(9295): p. 1793-1803.

13. Reshkin, S., et al., *Gonadotrophin releasing hormone (GnRH) receptor and steroid receptors in human uterine leiomyoma, myometrium and endometrium.* International Journal of Oncology, 1997. **11**(3): p. 603-607.

14. Tombal, B., *Prostate Cancer Management: What Does the Future Hold?* European Urology Supplements. **9**(8): p. 706-714.

15. Katsila, T., et al., *Evaluation of a Stable Gonadotropin-Releasing Hormone Analog in Mice for the Treatment of Endocrine Disorders and Prostate Cancer.* Journal of Pharmacology and Experimental Therapeutics. **336**(3): p. 613-623.

16. Chan, R.L. and M.D. Chaplin, *Plasma binding of LHRH and nafarelin acetate, a highly potent LHRH agonist.* Biochemical and Biophysical Research Communications, 1985. **127**(2): p. 673-679.

17. Nestor, J.J., et al., *Synthesis and biological activity of some very hydrophobic superagonist analogs of luteinizing hormone-releasing hormone.* Journal of Medicinal Chemistry, 1982. **25**(7): p. 795-801.

18. Padula, A.M., *GnRH analogues--agonists and antagonists.* Animal Reproduction Science, 2005. **88**(1-2): p. 115-126.

19. Bolton, E.E., et al., *Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities*, in *Annual Reports in Computational Chemistry*. 2008, Elsevier. p. 217-241.

20. Chemaxon, *MarvinSketch 5.3.1*, http://www.chemaxon.com/products/marvin/marvinsketch/.

21. Pappa, E.V., et al., *Enzymatic stability, solution structure, and antiproliferative effect on prostate cancer cells of leuprolide and new gonadotropin-releasing hormone peptide analogs.* Peptide Science, 2011. **96**(3): p. 260-272.

22. Millar, R.P., et al., *Diversity of actions of GnRHs mediated by ligand-induced selective signaling.* Frontiers in neuroendocrinology, 2008. **29**(1): p. 17-35.

23. Deslauriers, R., et al., *Conformational flexibility of luteinizing hormone-releasing hormone in aqueous solution. Carbon-13 spin-lattice relaxation time study.* Biochemistry, 1975. **14**(19): p. 4335-4343.

24. Freidinger, R.M., et al., *Bioactive conformation of luteinizing hormone-releasing hormone: evidence from a conformationally constrained analog.* Science, 1980. **210**(4470): p. 656-658.

25. Kubli-Garfias, C. and P.M. Conn, *Electronic structure of gonadotropin-releasing hormone (GnRH): a semiempirical study.* Journal of Molecular Structure: THEOCHEM, 2000. **529**(1-3): p. 203-208.

26. Guarnieri, F. and H. Weinstein, *Conformational Memories and the Exploration of Biologically Relevant Peptide Conformations:â€‰ An Illustration for the Gonadotropin-Releasing Hormone.* Journal of the American Chemical Society, 1996. **118**(24): p. 5580-5589.

27. Momany, F.A., *Conformational energy analysis of the molecule, luteinizing hormone-releasing hormone. 1. Native decapeptide.* Journal of the American Chemical Society, 1976. **98**(10): p. 2990-2996.

28. Marche, P., et al., *Conformational characteristics of luliberin. Circular dichroism and fluorescence studies.* Biochemistry, 1976. **15**(26): p. 5730-5737.

29. Milton, R.C.d.L., et al., *Comparative structure-activity studies on mammalian [Arg8] LH-RH and chicken [Gln8] LH-RH by fluorimetric titration.* Biochemical and Biophysical Research Communications, 1983. **111**(3): p. 1082-1088.

30. Pan, J.-T., et al., *Electrophysiological test of an amphiphilic β-structure in LHRH action.* Molecular and Cellular Endocrinology, 1986. **48**(2-3): p. 161-166.

31. Mabrey, S. and I.M. Klotz, *Conformation of gonadotropin releasing hormone.* Biochemistry, 1976. **15**(1): p. 234-242.

32. Baniak, E.L., et al., *Nuclear magnetic resonance analysis and conformational characterization of a cyclic decapeptide antagonist of gonadotropin-releasing hormone.* Biochemistry, 1987. **26**(9): p. 2642-2656.

33. Sprecher, R.F. and F.A. Momany, *On the conformation of Luteinizing Hormone-Releasing Hormone, nuclear Overhauser observation.* Biochemical and Biophysical Research Communications, 1979. **87**(1): p. 72-77.

34. Deslauriers, R., R. Walter, and I.C.P. Smith, *13C nuclear magnetic resonance studies of the conformation of the X-Pro bond in the oligopeptide hormones, thyrotropin-releasing hormone, luteinizing hormone-releasing factor, angiotensin and melanocyte-stimulating hormone release-inhibiting factor.* Biochemical and Biophysical Research Communications, 1973. **53**(1): p. 244-250.

35. Sawaya, M.R., et al., *Atomic structures of amyloid cross-β spines reveal varied steric zippers.* Nature, 2007. **447**(7143): p. 453-457.

36. Zimarino, V., S. Wilson, and C. Wu, *Antibody-mediated activation of Drosophila heat shock factor in vitro.* Science, 1990. **249**(4968): p. 546-549.

37. Makin, O.S., et al., *Molecular basis for amyloid fibril formation and stability.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(2): p. 315-320.

38. Guddat, L.W., et al., *Three-Dimensional Structure of the α-Conotoxin GI at 1.2 Å Resolution.* Biochemistry, 1996. **35**(35): p. 11329-11335.

39. Kochman, K., et al., *Binding of Cu2+, Zn2+, and Ni2+¸GnRH complexes with the rat pituitary receptor.* Journal of Inorganic Biochemistry, 1997. **65**(4): p. 277-279.

40. Winn, M.D., et al., *Overview of the CCP4 suite and current developments,* in *Acta Crystallographica Section D*. p. 235-242.

41. Kantardjieff, K.A. and B. Rupp, *Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals.* Protein Science, 2003. **12**(9): p. 1865-1871.

42. Allen, F., *The Cambridge Structural Database: a quarter of a million crystal structures and rising,* in *Acta Crystallographica Section B*2002. p. 380-388.

43. Benedetti, E., et al., *Linear oligopeptides. Part 227. X-Ray crystal and molecular structures of two α-helix-forming (Aib-L-Ala)sequential oligopeptides, pBrBz-(Aib-L-Ala)5-OMe and pBrBz-(Aib-L-Ala)6-OMe.* J. Chem. Soc., Perkin Trans. 2, , 1990: p. 1829-1837

44. Karle, I.L., et al., *Modular design of synthetic protein mimics. Characterization of the helical conformation of a 13-residue peptide in crystals.* Biochemistry, 1989. **28**(16): p. 6696-6701.

45. Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases.* Biopolymers, 2008. **89**(5): p. 392-400.

46. Zhong, L. and W. Johnson, *Environment affects amino acid preference for secondary structure.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4462-4465.

47. Shepherd, N.E., et al., *Single Turn Peptide Alpha Helices with Exceptional Stability in Water.* Journal of the American Chemical Society, 2005. **127**(9): p. 2974-2983.

48. Yamamoto, Y., et al., *Conformational requirement of signal sequences functioning in yeast: circular dichroism and proton nuclear magnetic resonance studies of synthetic peptides.* Biochemistry, 1990. **29**(38): p. 8998-9006.

49. Nelson, J.W. and N.R. Kallenbach, *Persistence of the .alpha.-helix stop signal in the S-peptide in trifluoroethanol solutions.* Biochemistry, 1989. **28**(12): p. 5256-5261.

50. Roccatano, D., et al., *Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: A molecular dynamics study,* 2002. p. 12179-12184.

51. Stratakis, C.A., et al., *Hypothalamic Hormones GnRH, TRH, GHRH, SRIF, CRH, and Dopamine*, in *Endocrinology*, S. Melmed and P.M. Conn, Editors. 2005, Humana Press: New Jersey. p. 173-195.

52. Maynard, A.J., G.J. Sharman, and M.S. Searle, *Origin of β-Hairpin Stability in Solution: Structural and Thermodynamic Analysis of the Folding of a Model Peptide Supports Hydrophobic Stabilization in Water.* Journal of the American Chemical Society, 1998. **120**(9): p. 1996-2007.

53. de Alba, E., M.A. Jiménez, and M. Rico, *Turn Residue Sequence Determines β-Hairpin Conformation in Designed Peptides.* Journal of the American Chemical Society, 1997. **119**(1): p. 175-183.

54. Haque, T.S. and S.H. Gellman, *Insights on Jiménez-Hairpin Stability in Aqueous Solution from Peptides with Enforced Type I and Type II β-Turns.* Journal of the American Chemical Society, 1997. **119**(9): p. 2303-2304.

Chapter 5. Identification of small molecule inhibitors to GnRH-R using similarity searching.

As previously discussed in chapter 1 there are currently no clinically approved small molecule antagonists of the GnRH receptor, meaning that all treatments are currently peptide based. Such therapy requires routine intramuscular injection or use of a slow release implant. These types of treatment are not only more inconvenient to the patient but they are also expensive to produce. There is therefore strong interest in developing small molecule antagonists; some progress has been made in this direction.

5.1 Elagolix is a non-peptide orally active GnRH-R antagonist.

During the course of this study the first orally active small molecule GnRH antagonist was found and entered into drug trials. Sodium R-(+)-4-{2-[5-(2-fluoro-3-methoxyphenyl)-3-(2-fluoro-6-[trifluoromethyl]benzyl)-4-methyl-2,6-dioxo-3,6-dihydro-2H-pyrimidin-1-yl]-1-phenylethylamino}butyrate (branded Elagolix)[1] is a selective hGnRH receptor binding molecule with a dissociation constant of 54 pM[2]. The drug is expected to enter phase III clinical trials this year.

5.1.1 Elagolix development through structure activity relationships.

The synthesis of the Elagolix structure (Figure 5-1) was a result of chemical addition to a high affinity thienopyrimidinedione scaffold first identified in 1998[3]. Unfortunately this scaffold also had a high affinity for the CYP3A4 protein. The authors found that certain modifications not only increased the affinity for the GnRH receptor but also decreased the affinity to CYP3A4 to a far more tolerable level[1].

Looking at Figure 5-2 the design decisions taken to produce the final molecular structure can be understood.

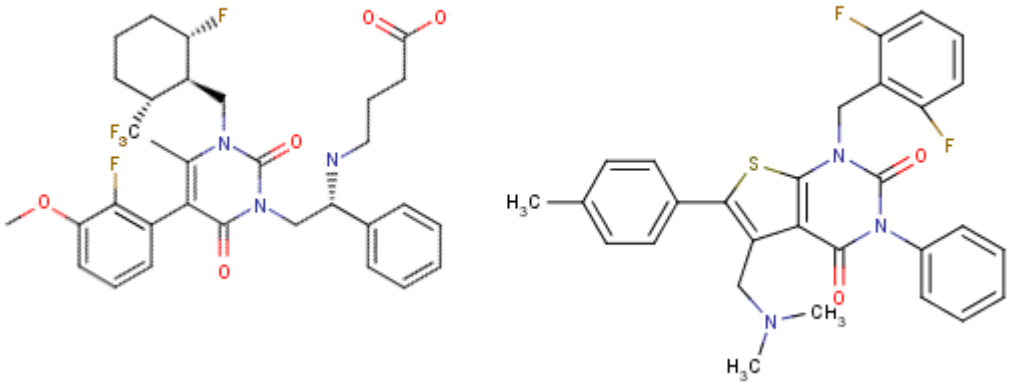


Figure 5-1. 2D structure of the orally active non-peptide antagonist Elagolix (left) and the thienopyrimidinedione scaffold that it was built on (right).

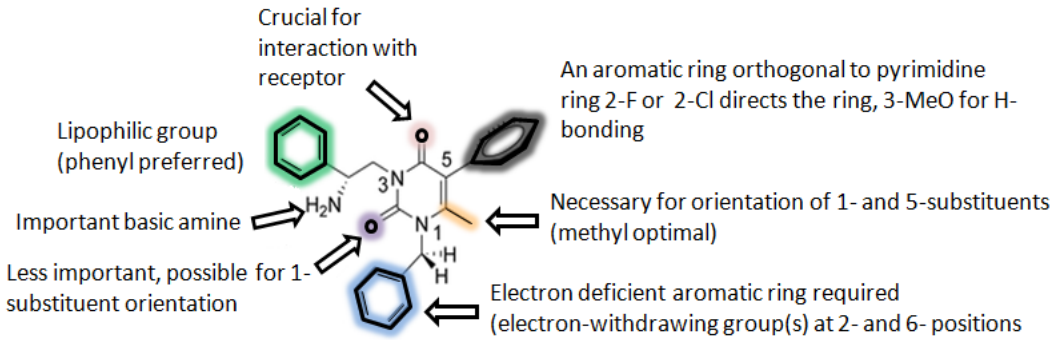


Figure 5-2. The uracil pharmacaphore used in Elagolix has been designed for high affinity. Modified from reference [1].

5.1.2 Docking the Elagolix molecule into the GnRH Receptor

Whilst the binding interaction of thienopyrimidinedione antagonists has been studied previously this was expanded upon to focus specifically on the possible interactions between Elagolix with the receptor. Of particular interest is how additions to the

scaffold change binding mechanics. Using AutoDock vina the possible alternative binding modes of Elagolix are explored and compared against results of docking the thienopyrimidinedione scaffold into the β_2 adrenergic receptor model. Since the scoring system between docking software is so varied[4] the accuracy of the results is also an interesting point of discussion.

5.1.3 Thienopyrimidinedione binding site analysis

The interaction between the thienopyrimidinedione scaffold used in the Elagolix molecule and GnRH-R has been probed[5]. In a previous study it was proposed that the scaffold made 2 key hydrogen bonds with D302 and H306 (Figure 5-3). Only one of these hydrogen bonds extends to one of the key residues that bind GnRH described earlier in Figure 4-9. The same scaffold was docked using AutoDock vina into the β_2AR receptor model created in this work and the binding modes compared in Figure 5-3 and Figure 5-5.

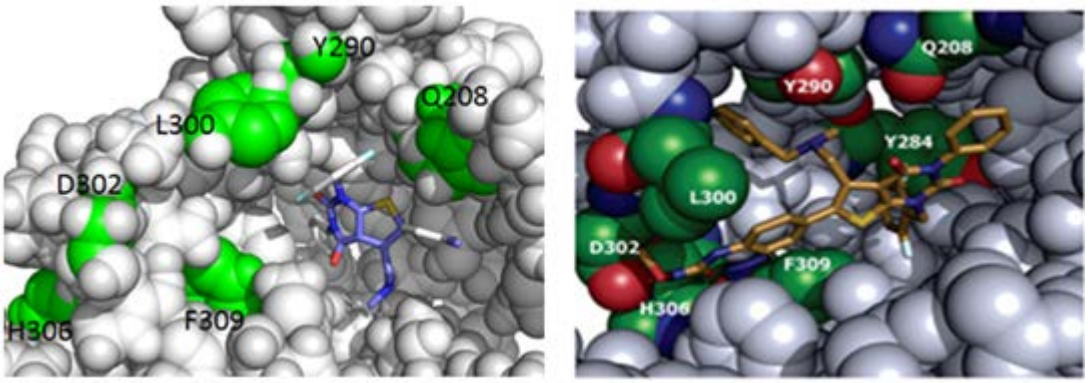


Figure 5-3. Comparison of the suggested thienopyrimidinedione scaffold binding in the β2AR derived receptor model using AutoDock vina (left) against the literature docking study from Betz et al[5] (right).

The docked scaffold resides in very different positions between both models, in the β_2AR derived model the scaffold is positioned into the slightly deeper pocket closer to Q208 than was found in the literature study. There are also distinct differences

between model structures in the positioning of the highlighted amino acids in Figure 5-3. The key difference is the lack of binding pocket formed by Y290, L300 and F309 shown in the literature docking, this cannot form in our β_2AR model because the L300 residue is on the opposite side of the helix.

5.1.4 Docking Elagolix using AutoDock vina

The Elagolix structure itself was docked it into the β_2AR receptor model using AutoDock vina and the most favourable pose is shown in Figure 5-4. We compared how the ligand would be posed in comparison to the more time consuming approach favoured by the authors, where they performed a series of molecular dynamics annealing experiments at various temperatures in a simulated solvated state before final energy minimisation was applied[5].

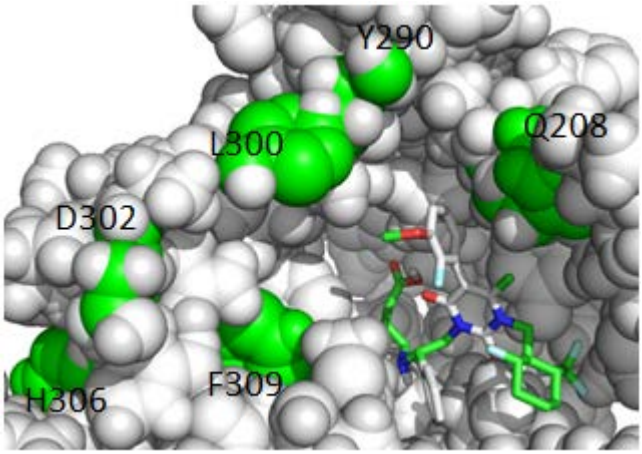


Figure 5-4. Docking Elagolix into the β_2 Adrenergic receptor derived model using AutoDock vina shows that as in the scaffold docking experiment the elagolix molecule is positioned into the deeper pocket closer to Q208.

5.1.5 Comparing docked poses of thienopyrimidinedione scaffold and Elagolix

If the β_2 AR and published models are compared there are key differences in each model with regards to where the highlighted residues are placed. Atomic coordinates for the published model are not available to align against our model. The published scaffold binding pose (Figure 5-4) shows a rather different binding pose to that obtained in our work (Figure 5-3), though both structures make reasonable hydrogen bonds with the receptor (Figure 5-5). The scaffold affinity for our receptor model was -9.4 kcal /mol which is higher than the score of -8.8kcal / mol obtained for elagolix. This is interesting since the scaffold shows only two possible hydrogen bonds compared to the three bonds that Elagolix can form with Tyr284 and Asn212 (Figure 5-5).

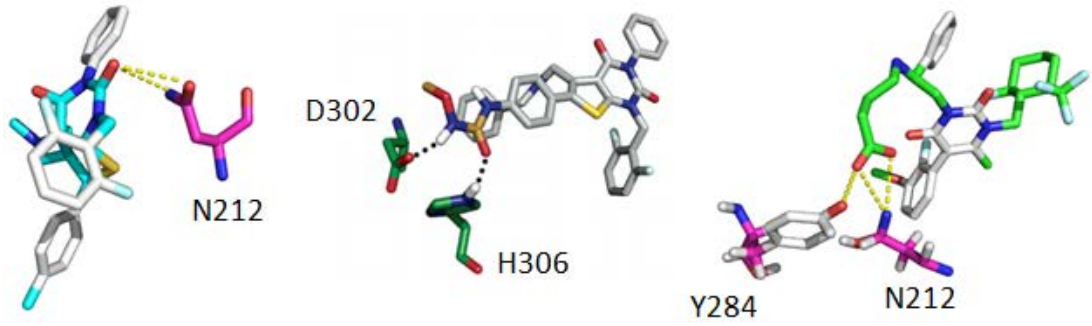


Figure 5-5. Comparison of docked binding poses between thienopyrimidinedione scaffold from this study (left) and the literature (middle) and Elagolix in the GnRH-R active site (right).

5.1.6 Probing the antagonist activity of Elagolix

It is not yet known if Elagolix acts allosterically and causes a structural change in GnRH-R which prevents GnRH association, or if Elagolix acts as a competitive inhibitor to prevent GnRH-I entry. To test the hypothesis that Elagolix acts as a

competitive inhibitor both the GnRH-I peptide and the Elagolix molecule were docked into the same receptor model to determine how closely the molecules would be if both occupied the active site simultaneously (Figure 5-6).

From this docking data it can be observed that the Elagolix molecule occupies a significant area of the peptide binding site which would occlude the peptide from being able to gain entry, particularly given the far tighter binding affinity of the small molecule over the peptide sequence.

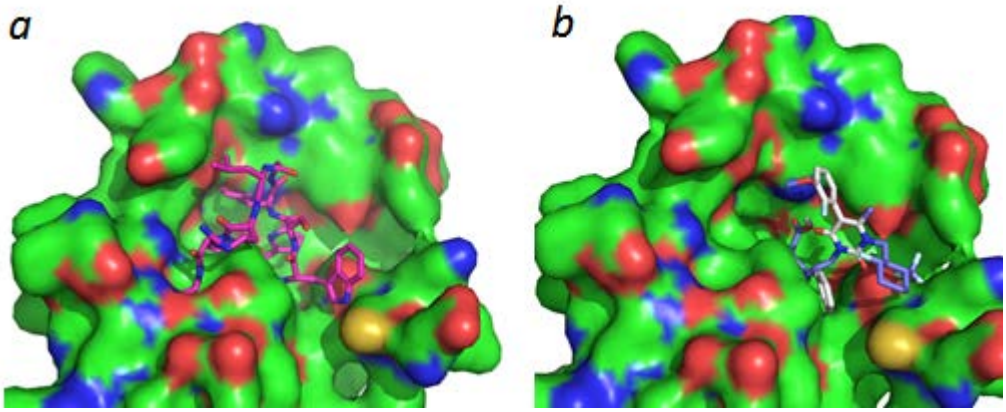


Figure 5-6. Comparison of the docked poses of the GnRH-I peptide (*a*) and Elagolix (*b*) it can be seen that Elagolix occupies almost half of the space of the GnRH-I peptide.

Closer inspection of the Elagolix docking suggests that it does not act as a GnRH-I peptide mimic; whilst the Elagolix molecule occupies similar space it does not make the same contacts as the peptide. Mimicry of the hairpin was a feature of the original scaffold.

Whether active site occupation alone is the only factor affecting GnRH binding is not clear however there are also likely to be small conformational changes to the receptor structure.

The Elagolix structure is a small molecule antagonist with a high affinity for the GnRH receptor. As a technique in developing further small molecule antagonists the Elagolix structure was used to search the EDULISS database for similar compounds.

5.2 Using the EDinburgh University LIgand Selection System (EDULISS) as a search space for compound similarity.

As described in Chapter 2.6 EDULISS is a database of chemical structures with associated descriptors. In this section the use of EDULISS as a search space for finding structures similar to Elagolix will be investigated. Similar compounds that are identified are then docked into the β_2AR model using AutoDock vina and scored. Any compounds that are identified as suitable candidates from docking are then assayed for activity.

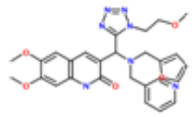
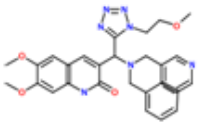
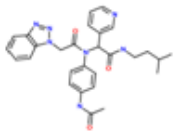
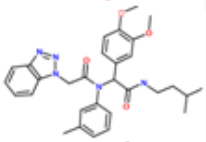
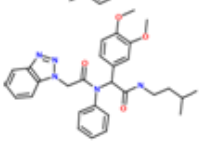
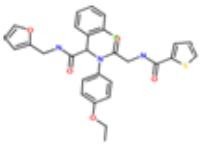
5.3 Searching EDULISS for compounds similar to Elagolix using Ultra Fast Shape Recognition with Atom Type (UFSRAT) and docking with AutoDock vina

Taking the structure of Elagolix (Neurocrine Biosciences) as a lead template a search of the EDULISS database was carried out using the software program UFSRAT as described in Chapter 2.8 to suggest compounds from the library which share similar structural characteristics (and therefore the possibility of similar interactions) with Elagolix.

The top 100 compounds had a similarity range of between 0.95 and 0.92, six were selected for assay (Table 5-1). These compounds had a UFSRAT similarity from

0.95 to 0.93 and were docked and scored using Autodock vina as described in Chapter 2.5. The identity searched compounds had a predicted affinity range of -9.0 to -8.1 kcal / mol; docking the Elagolix structure gives a score of -8.8 kcal / mol. The similar scoring to the Elagolix molecule is a good indicator that the compounds will have some form of activity.

Table 5-1. The structure of Elagolix was run through UFSRAT and the top 10 similar compounds were scored in AutoDock vina for predicted affinity.

Structure	Compound number	Asinex reference	Similarity score	Predicted affinity (kcal/Mol)	Hydrogen bonds to	Inhibition in 100μM IP3 assay
	14	ASN 04423417	0.95	-8.1	Ala209, Asn212, Ser118	36%
	15	ASN 04423498	0.94	-8.2	Tyr283, Asn212, Ser118	62%
	16	ASN 05296677	0.93	-8.3	Ser118	36%
	17	ASN 05296481	0.93	-8.5	Asn212	62%
	18	ASN 05296489	0.93	-8.1	Asn212	46%
	19	ASN 05990009	0.93	-9.0	Lys121, Ala209, Asn212, Tyr211	49%

The docked poses of the compounds were then compared to that of Elagolix (Figure 5-7). It was surprising to see compound 19 having a predicted affinity higher than Elagolix given that the Elagolix structure is so well engineered. However we can see from the docking results that this is due to the compounds ability to make four

hydrogen bonds to the receptor instead of the three hydrogen bonds of elagolix. All compounds were docked into the same pocket as Elagolix with 1-3 potential hydrogen bonds. Hydrogen bonding to Asn212 was common to all compounds except compound 16 which has a hydrogen bond to Ser118.

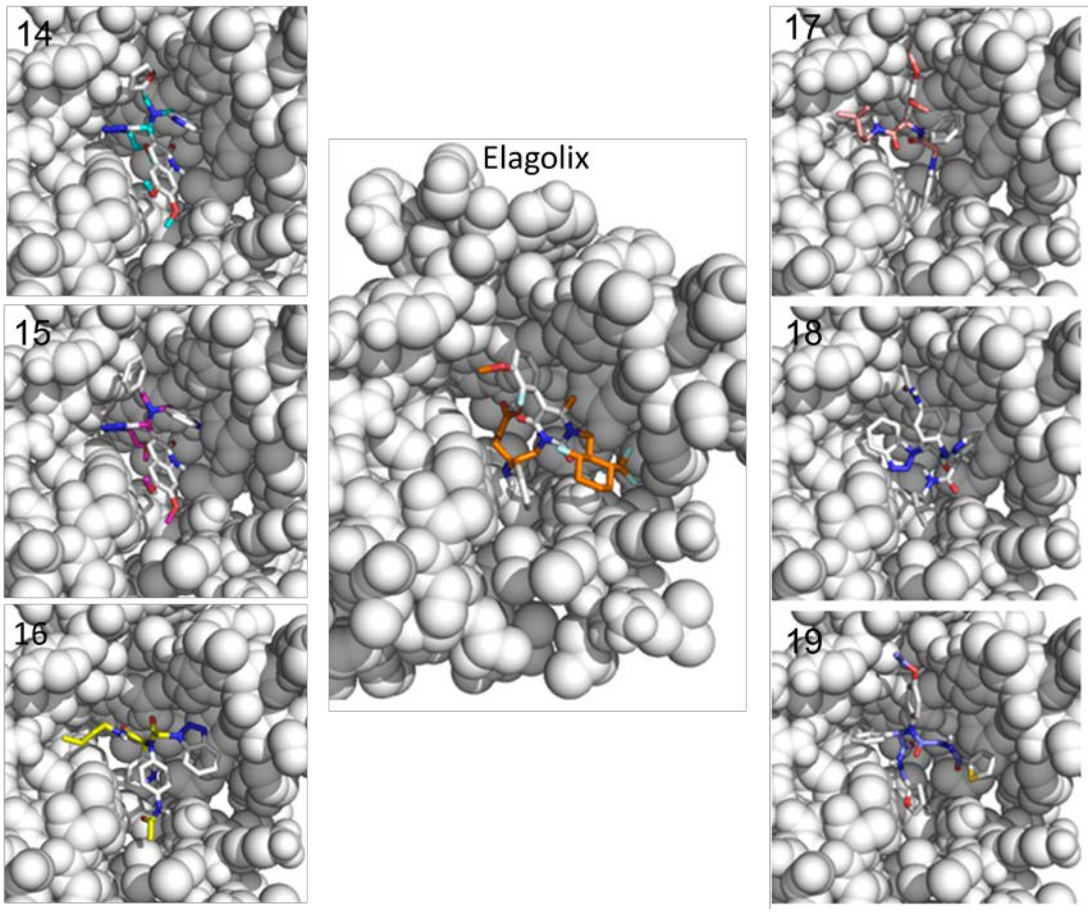


Figure 5-7. Comparisons of binding locations between compounds found from EDULISS with high similarity to Elagolix with Elagolix. Compound numbers are shown in the upper left corner of each pose.

5.4 Assaying compounds structurally similar to Elagolix with the GnRH receptor.

Six compounds (compounds 14-19) which were readily available from Asinex were selected for assay. Initially a single point IP assay at 100 µM was performed to quickly assess if any of the compounds showed antagonist activity (Table 5-1). The assay mechanism relies on the activated signalling pathway of GnRH-R using IP3 as a second messenger. As tritiated inositol is added to the media prior to assay, the activation of GnRH-R can be followed by lysing the cells and measuring radioactivity of IP3 bound to resin. Compounds which are able to prevent GnRH-R activation by GnRH-I peptide will show lower radioactivity than the GnRH-I stimulated reference cells. The technical details are described in Chapter 2.10.4. The difference in inhibition between compounds 14 and 15 was particularly interesting given that they only differ by a single oxygen atom. Their binding poses were also very similar, however compound 14 made a hydrogen bond to Ala209 instead of Tyr283. From this screen compounds 14, 15, 17 and 19 were used in dose-response competitive binding assay.

5.4.1 Whole cell binding assay

The whole cell binding assay was performed as described in Chapter 2. Briefly this assay measures the decrease in radioactivity from competition binding. Competition arises between increasing concentrations of the test compound and an ^{125}I-[His^{5}, D-Tyr^{6}] –GnRH tracer peptide of high affinity. Therefore the lower the radioactivity reading, the better the compound is as a competitive inhibitor.

Results in Figure 5-8 suggest that all compounds were able to compete off the radiolabelled peptide to some degree. As such the compounds were then tested in a dose –response assay (details in Chapter 2) to measure inhibition of inositol 1,4,5,-trisphosphate (IP3) production (Figure 5-9).

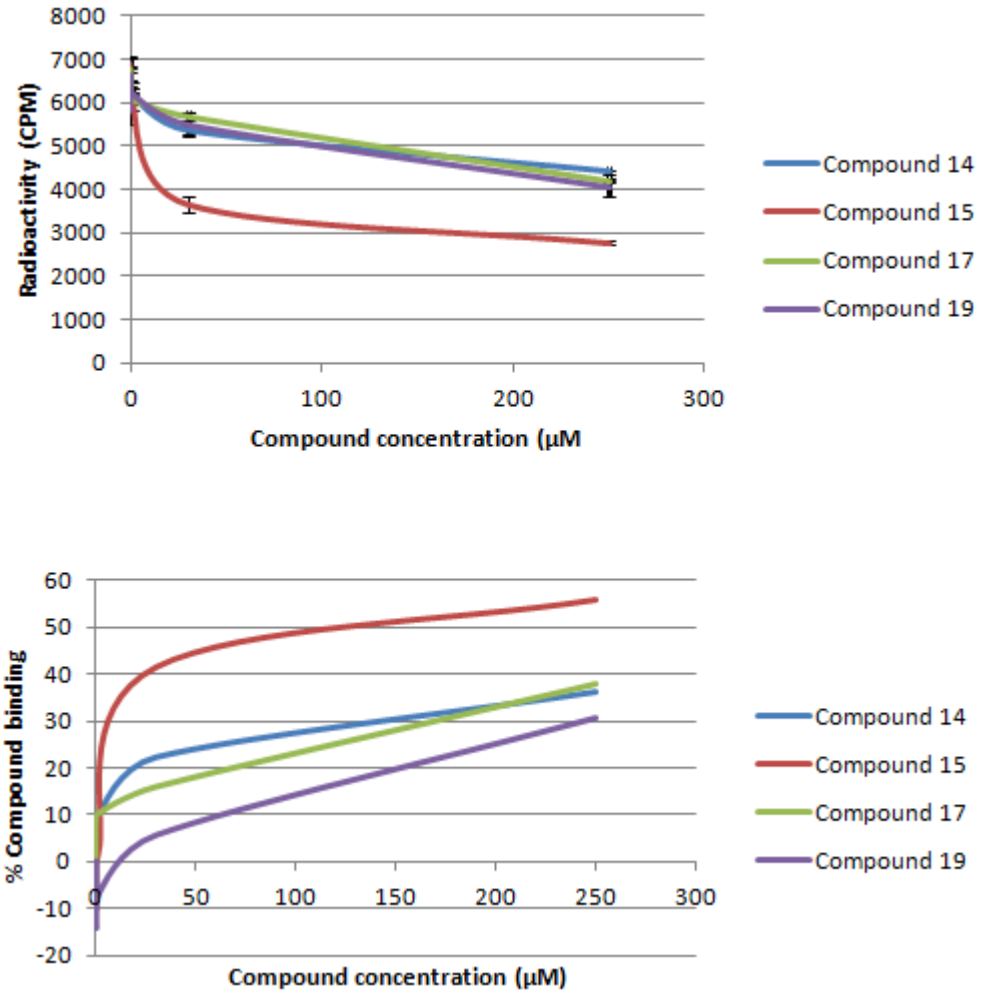


Figure 5-8. Results of whole cell binding assay of elagolix similar compounds at 0.01, 0.1, 1.0, 30 and 250µM concentrations. (Top) Raw bound radioactivity reading. (Bottom) Radioactivity reading expressed as a percentage of bound compound.

5.4.2 Inositol Phosphate inhibition (IP3) Assay

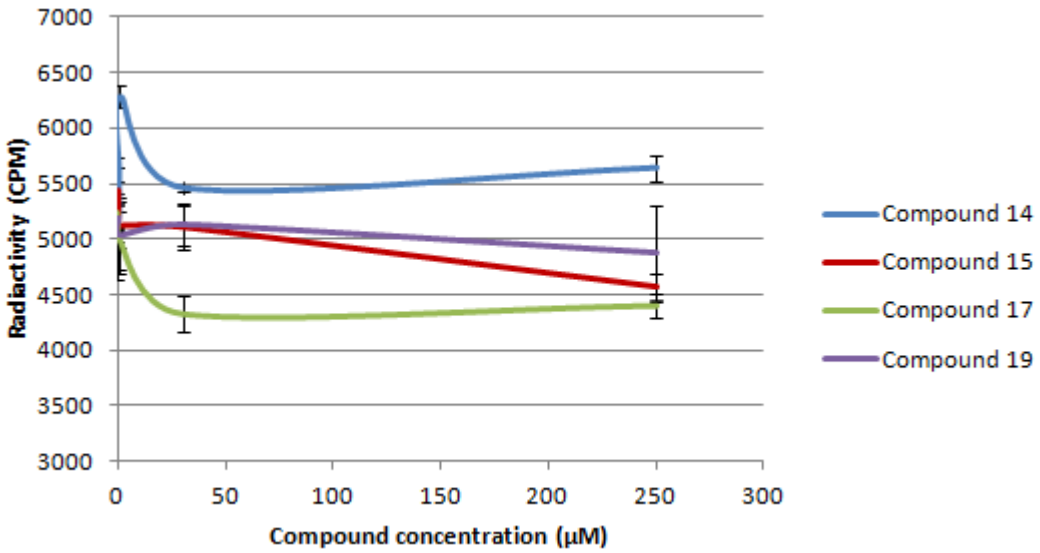


Figure 5-9. Radioactivity from deuterated inositol produced by stimulation with GnRH peptide. Compounds are loaded onto cells prior to activation with agonist GnRH peptide and inhibition of IP3 production is used to measure how well the test compound is able to occlude the peptide from GnRH-R activation.

In contrast to the whole cell binding results the IP3 assay showed that even at the highest compound concentration there was little receptor activation inhibition.

5.4.3 Discussion of assay results for compounds selected by Elagolix similarity

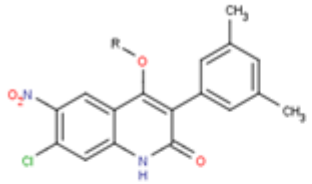
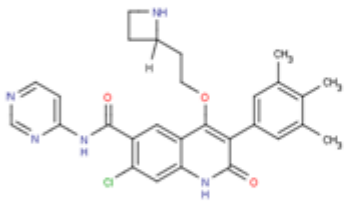
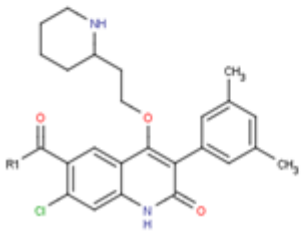
Figure 5-8 showed a fairly strong result for compound 15. 40% of the compound had bound to the receptor at the 30 µM point. However the same effect was not observed in the IP3 assay. One reason for this may be that the compound was displacing far more non-specifically bound radio labelled peptide than that bound to the active site. The results from the IP3 assay with compound 14 and 19 are also interesting to examine; after initially causing an inhibition response at 30 µM there was no further effect from increasing concentration despite showing increased radioligand displacement up to 250 µM in the binding assay. This is possibly due to the effect of

non-specific binding of the radioligand which is being displaced by the compound but which has no effect on IP3 production.

5.5 Quinolone compounds as potential GnRH-R antagonists

There have been a number of reports in the literature[6-9] which have identified quinolone (Table 5-2) based compounds as being antagonists to the GnRH receptor at high affinity. As a means to better understand the QSAR of these compounds and to further develop the scaffold, a select number of quinolone and quinolone-like compounds from a focussed inhouse compound library collection were tested to probe for antagonist activity.

Table 5-2. IC50 values and structures for quinolone GnRH-R inhibitors found in the literature.

Structure	IC50 (nM)	Reference
	10	DeVita 1999
	0.44	DeVita 2004
	0.8	Walsh 1999

5.5.1 Selection of quinolone/quinolone-like compounds and AutoDock vina scoring

In total 11 compounds from the inhouse quinolone-like library were docked and scored (Table 5-3), these were selected on the basis of the presence of the quinolone group or where the group was absent sufficient quinolone likeness. Solubility data provided by the manufacturer was also a consideration.

5.5.2 Assay of quinolone compounds

Due to limited materials it was only possible to assay one concentration of each compound in a radioligand binding assay. All compounds were assayed at 33 µM except compound A10 which was 66 µM, the GnRH peptide control was 1 nM.

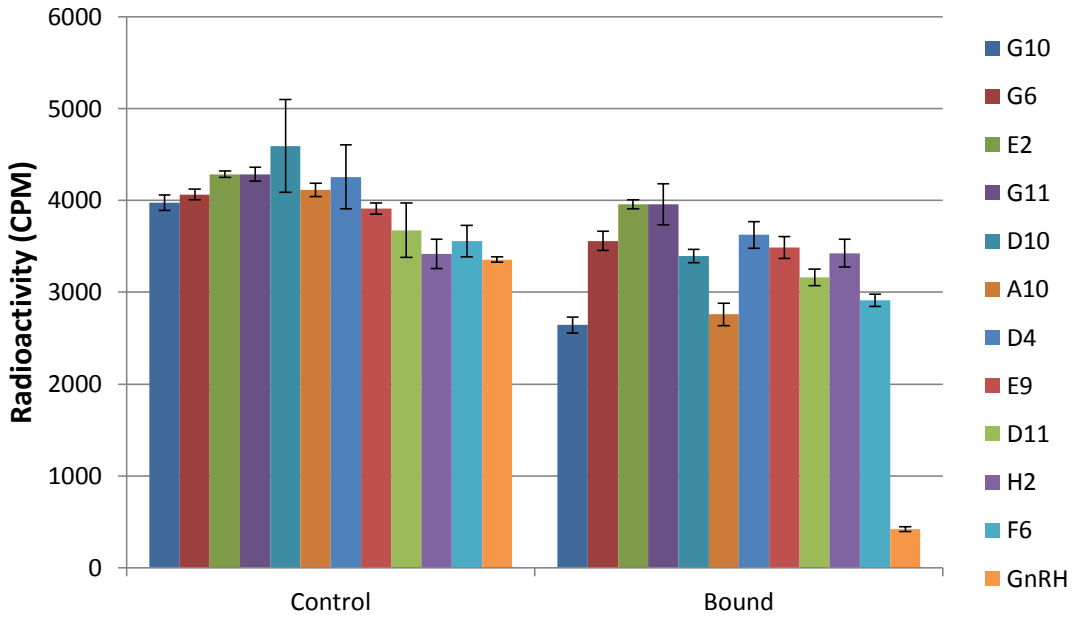
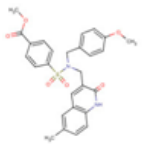
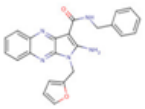
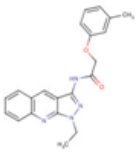
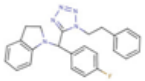
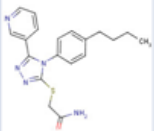
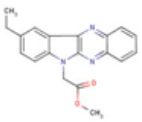
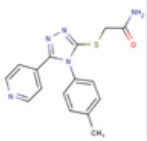
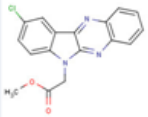
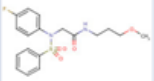
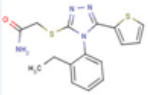
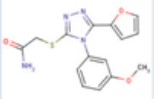


Figure 5-10. Single point assay results for quinolone/quinolone like compounds. Results shown are for 33µM of the respective compound excluding A10 (66µM) and GnRH (1nM). All conditions include 5% DMSO.

Table 5-3. Selected quinolone/quinolone like compounds assayed for antagonist activity.

Compound code	Structure	Asinex code	Autodock affinity (kcal/mol)
D4		ASN 04191900	-9.2
D11		BAS 09528296	-9.2
G11		BAS 12533563	-8.6
A10		ASN 05834079	-8.5
G6		ASN 04692048	-7.9
D10		BAS 07100896	-7.7
E2		BAS 02054521	-7.6
G10		BAS 07403636	-7.6
E9		BAS 02205678	-7.3
F6		ASN 04689472	-7.3
H2		ASN 04792697	-7.3

5.5.3 Discussion on the potential of quinolone compounds

The assay results for quinolone and quinolone like compounds gave positive results. Of the eleven compounds assayed seven gave a decrease in bound radioligand readings. Interestingly the best scoring compounds did not show the largest binding activity. The largest change was 33% using compound G10 followed by 26% by compound D10. When docked into the receptor (Figure 5-11) these molecules do bind quite deeply into the active site. The position of D10 does not allow the formation of hydrogen bonds directly to any of the receptor residues however there may be a ring stacking interaction with Trp101. G10 is able to make two hydrogen bonds to residues Tyr211 and Asn212.

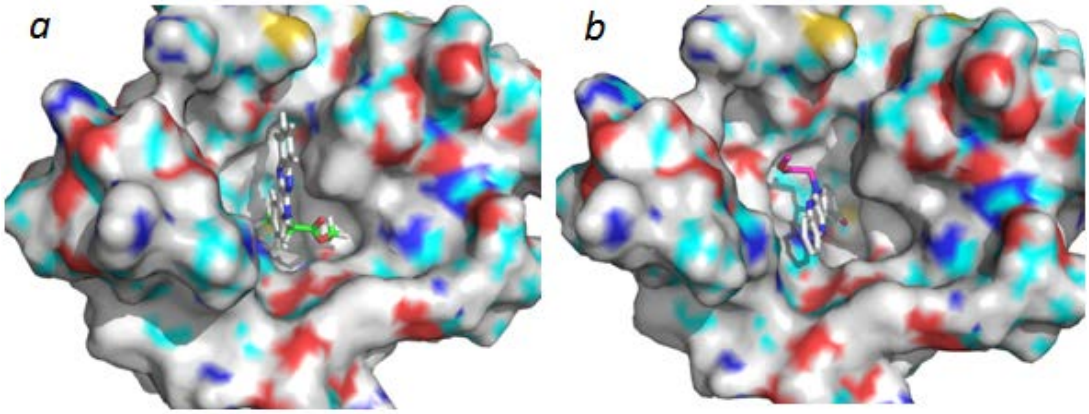


Figure 5-11. Docked poses of compounds D10 (b) and G10 (a). The compound positions are located deeply into the pocket in the centre of the active site and represent a good target location for antagonist activity.

5.6 Docking beta blockers into the "β"$_2$ adrenergic receptor derived GnRH receptor model demonstrates bias for β Adrenergic receptor ligands.

As a control experiment, three common beta blockers were docked into the GnRH receptor model to assess how the use of the β_2 Adrenergic receptor as a template structure affected the possibility of beta blocker structures scoring systematically high. There is no literature evidence indicating that beta blockers can bind to GnRH-R therefore this experiment aims to determine two points; do beta blockers score highly in docking experiments with GnRH-R and secondly, does this translate to inhibition activity in an IP3 assay?

5.6.1 Comparison of β adrenergic receptors.

There are three known beta adrenergic receptors transcribed in the human genome. As previously described they are all members of the rhodopsin like GPCR family of which GnRH-R is also a member. Beta blockers have been developed to be selective for the receptor type that they target, β_1, β_2 and non-selective blockers are all available. As the GnRH-R model that we produced was based on the β_2 adrenergic receptor, the structure of the model is heavily influenced by the helix conformations of β_2AR. This would make testing a β_2 selective drug a logical decision. However if the other β adrenergic receptors share high identity to the β_2AR then the drugs which target those receptors may also be useful. An alignment of the β adrenergic receptors was produced to establish this (Table 5-4 and Figure 5-12).

Table 5-4. Identity comparisons between the Beta Adrenergic receptor sequences.

Sequence 1	Length (AA)	Sequence 2	Length (AA)	Identity (%)
Beta 1 Receptor	477	Beta 2 Receptor	413	48.0
Beta 1 Receptor	477	Beta 3 Receptor	408	50.0
Beta 2 Receptor	413	Beta 3 Receptor	408	39.0

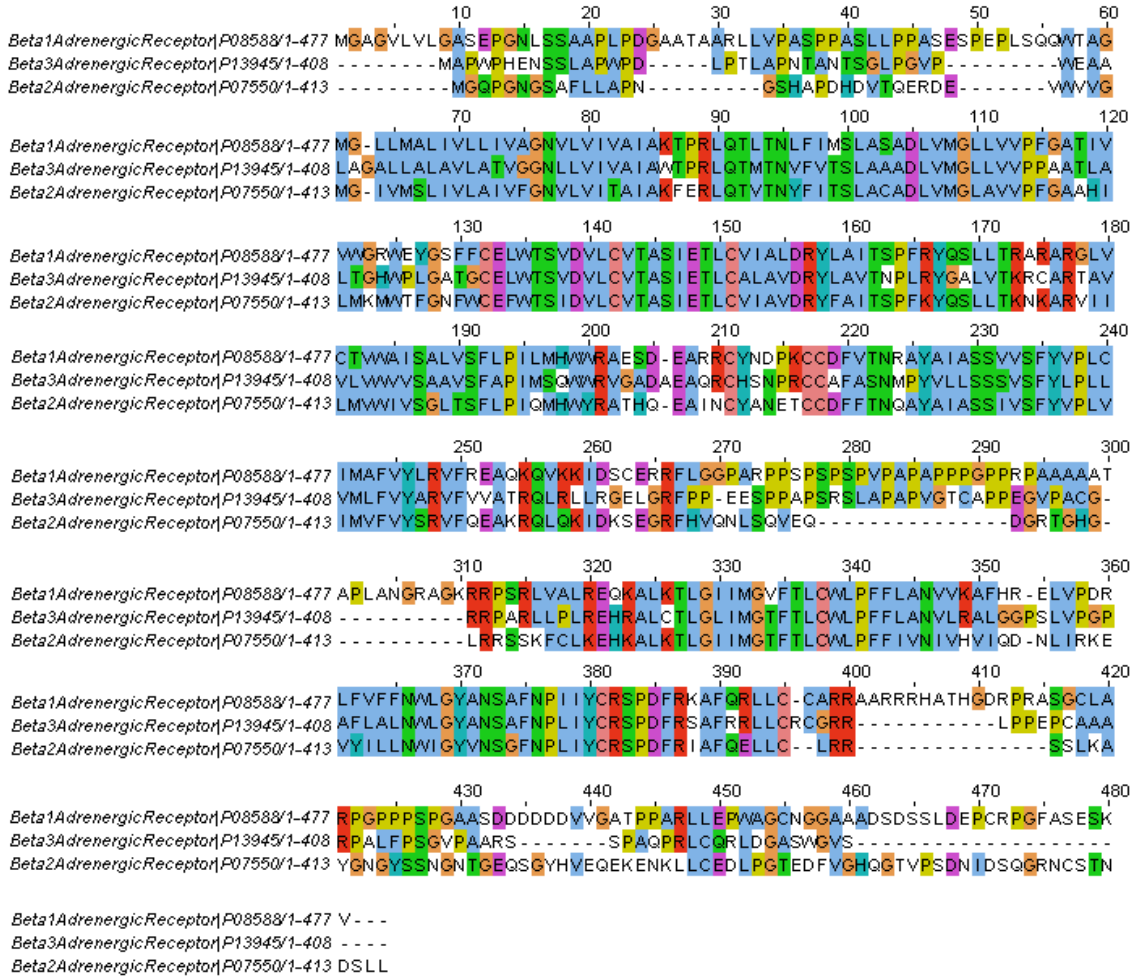


Figure 5-12. Sequence alignment of beta adrenergic receptors 1-3 indicates most homology is found in the membrane-bound regions of the receptor structures. The type 1 receptor shares 50% identity with the type 2 and 3 receptors. It was decided to include type 1 selective drugs in the docking simulations.

The three beta adrenergic receptors share between 39-50% identity. This is a fairly high score given that these are GPCR proteins. As observed previously the membrane bound regions of the proteins are the regions which share the most identity. As the identity scores were fairly high it is useful to dock ligand molecules for all receptor types against the GnRH receptor model that has been derived. After the model and subsequent docking experiments in this study were concluded using the β_2 receptor as a template, there was a subsequent publication of the β_1 adrenergic receptor. There is scope for the creation of a GnRH-R model using this structure. However given the high identity of the buried residues between types β_1 and β_2 the usefulness was questionable, especially when we consider the transmembrane regions of each protein can be superimposed with an RMSD of 0.809Å (Figure 5-13).

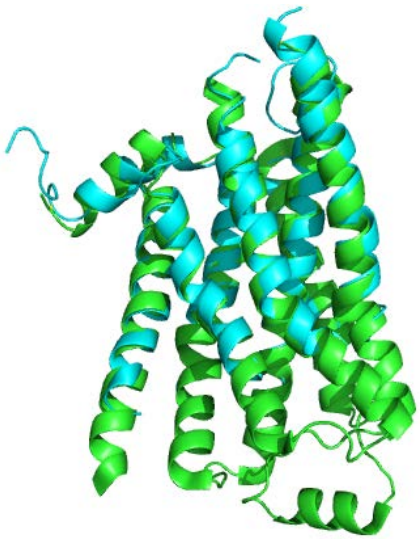
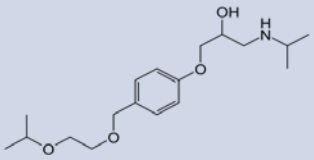
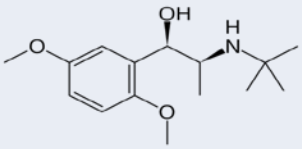
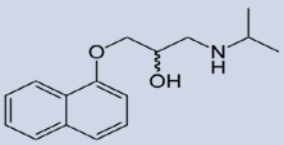


Figure 5-13. Superimposing the Turkey β_1 adrenergic receptor (PDB code 2VT4) and the Human β_2 adrenergic receptor shows an RMSD value of 0.809Å. Alignment of the protein sequences between the Turkey and Human β_1 adrenergic receptor shows 284 identical residues and 90 similar residues over the span of the ~480 amino acid sequence.

In the docking experiment we used three beta blocker structures; Bisoprolol (B1 selective), Butaxamine (B2 selective) and Propranolol (non-selective) to cover the full spectrum of drug selectivity for the adrenergic receptor (Table 5-5). The 3D structure generation and docking procedures were the same as described in Chapter 2.

Table 5-5. Beta blocker compounds used to assess bias in the GnRH-R model structure.

Compound Name	Adrenergic Receptor Selectivity	2D Structure
Bisoprolol	B1	
Butaxamine	B2	
Propranolol	No selectivity	

5.6.2 Docking the B1 Selective beta blocker Bisoprolol into the GnRH receptor.

The first docking result from the beta blocker family gave a consistently lower score for the GnRH receptor than both the peptide and non-peptide antagonists that have been docked thus far. The shape of the Bisoprolol molecule did however allow the

molecule to dock into the receptor in an elegant manner despite the lower affinity, creating potential hydrogen bonding to Asn305 (Figure 5-14).

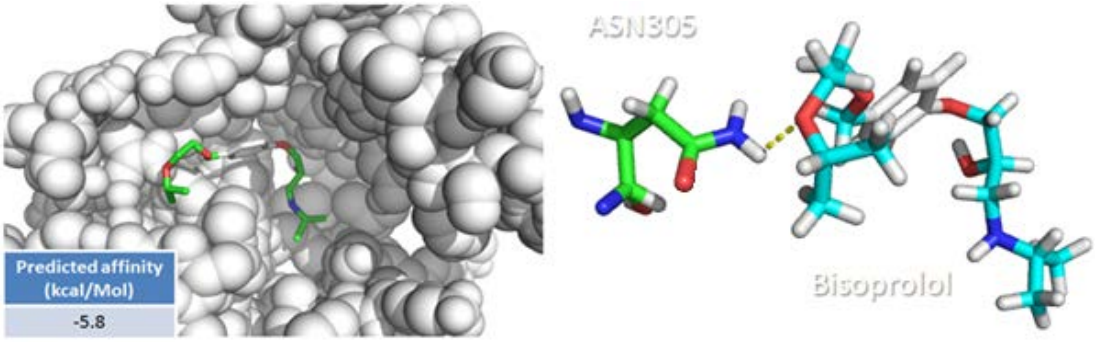


Figure 5-14. Docking the Bisoprolol molecule into the GnRH receptor indicated that one hydrogen bond could be formed between the ligand and receptor.

5.6.3 Docking the B2 Selective beta blocker Butaxamine into the GnRH receptor model.

The docked butaxamine molecule may form two hydrogen bonds with the receptor (Figure 5-15), this would involve residue Asn212. Interestingly although this interaction contains an extra hydrogen bond, the predicted affinity is the same as bisoprolol.

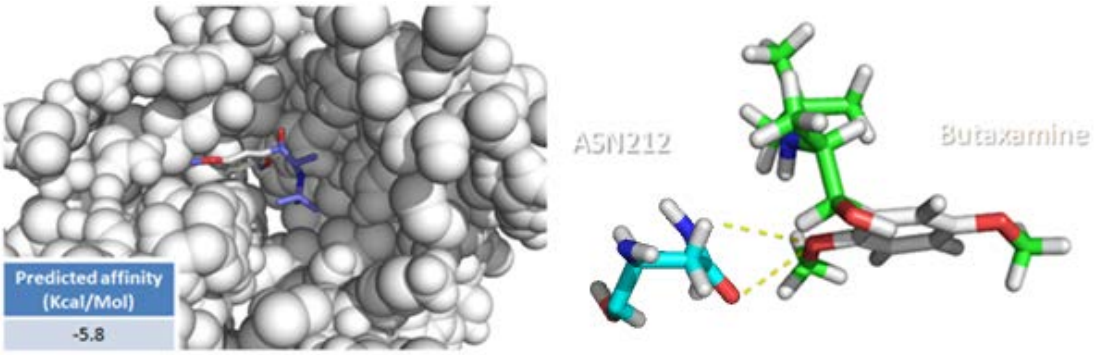


Figure 5-15. Docking the blocker Butaxamine into the GnRH receptor suggests hydrogen bonding to Asn212.

5.6.4 Docking the non-Selective beta blocker Propranolol into the GnRH receptor model.

The final beta blocker docking experiment produced the highest affinity of all molecules tested (Figure 5-16). This may be a result of the unselective properties of the ligand for a specific beta adrenergic receptor. In this case only potential hydrogen bonding to Leu286 seemed likely. The reason for increased affinity is unclear; initially as the ligand has an extra ring in the structure it was thought this might be able to form a ring stacking interaction but close scrutiny of the docked model indicates this not to be the case. It is noteworthy that the hydrogen bond location is however closer than in the other docking experiments.

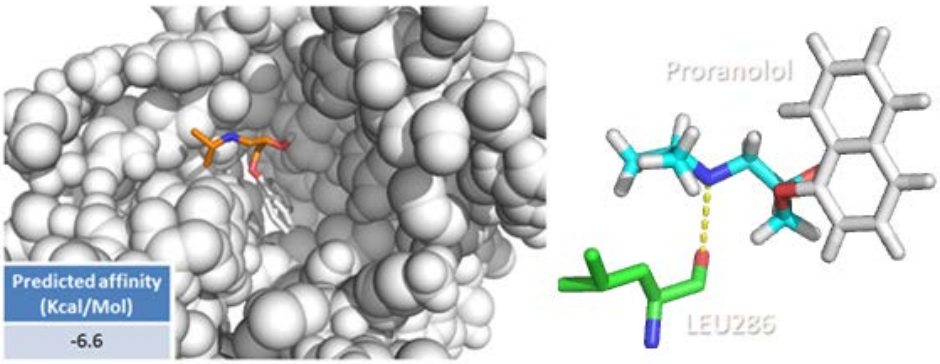


Figure 5-16. Docking the unselective beta blocker Propranolol into the GnRH receptor suggests hydrogen bonding to Leu286.

5.6.5 Inositol phosphate production assay indicates no GnRH receptor agonist or antagonist activity from Propranolol.

As the Propranolol molecule was the highest scoring compound from the virtual screening study the compound was assayed by measuring the effect that it had on inositol phosphate production at a concentration of 250 µM.

Reviewing the results of the inositol phosphate assay (Figure 5-17) indicates that no definitive increase or decrease in inositol phosphate was recorded. This indicates that 250 µM Propanolol has no binding interaction with the GnRH receptor regardless of how well it performed in the docking experiments.

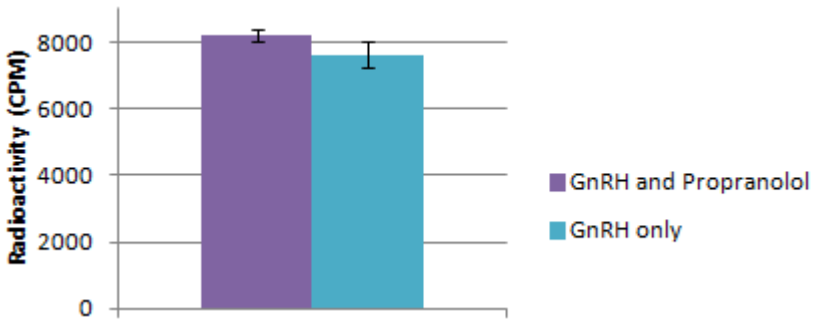


Figure 5-17. Comparison of Inositol phosphate production between the basal state and the 250µM Propranolol incubated cells shows that IP production is not significantly affected.

5.6.6 Homology models have a degree of bias for the template structure ligands

The first aim of this study was to establish if beta blockers scored highly in docking experiments using a model derived from a beta adrenergic receptor. It was found that all beta blockers docked with poorer scores than GnRH-R specific ligands. To further explore this idea the same three beta blockers were docked into the inactive rhodopsin derived GnRH-R model. All three beta blockers scored an affinity

approximately 1 kcal/ mol less than was observed for the beta adrenergic derived GnRH-R model (Figure 5-18).

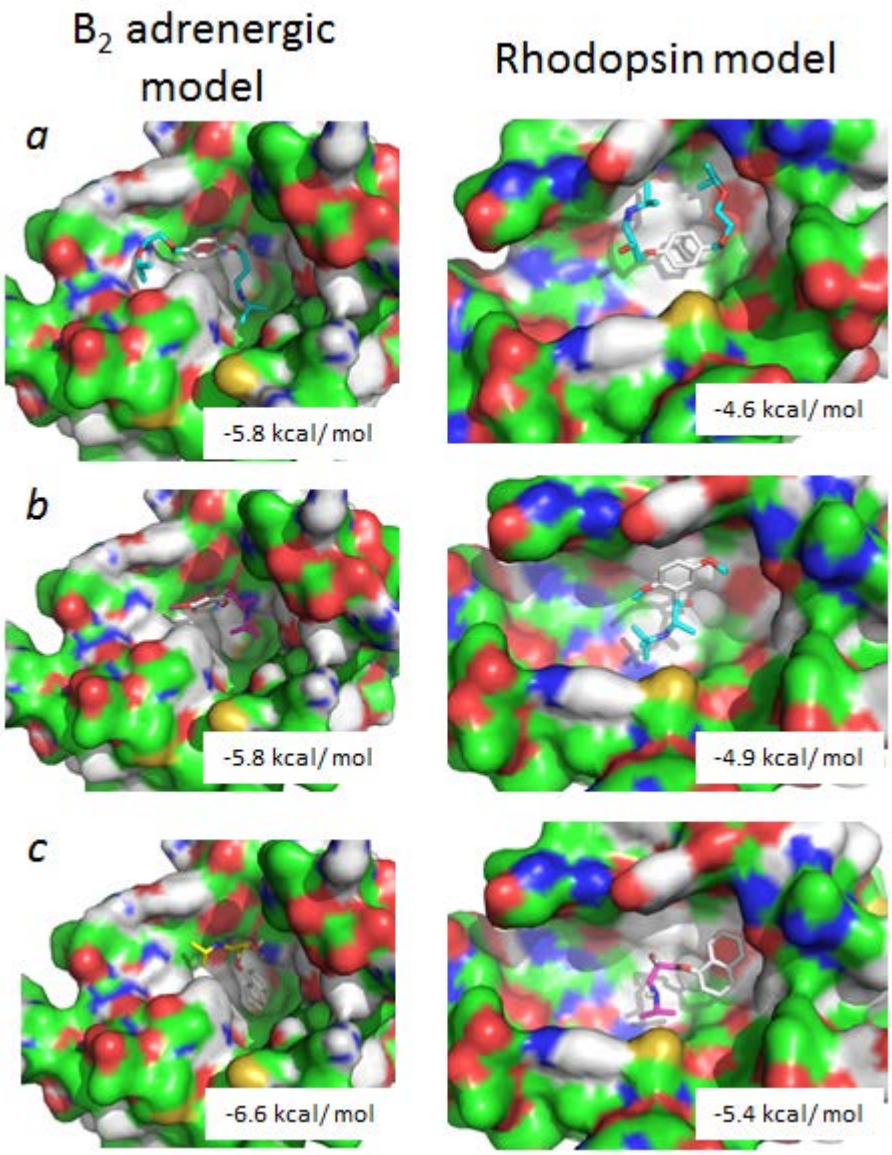


Figure 5-18. Comparison of docked beta blocker positions and scores between GnRH-R models derived from the β_2-adrenergic receptor and rhodopsin shows higher scoring in the β_2-adrenergic receptor model.

It is interesting that the non-selective beta blocker scored the most favourably in both models, this may be because the fused ring structure allows the molecule to make more favourable contacts than the selective beta blockers.

Assay results showed that despite the docking results being favourable there was no measurable interaction between the beta blocker and receptor. Therefore when doing virtual screening using a model structure it is useful to examine top hits for similarities to ligands for the template structure so that they can be assessed more rigorously prior to purchasing for assaying.

5.7 Conclusions and future work

The main goal of this chapter was to address how similarity to a known small molecule inhibitor of GnRH-R could be exploited to data mine similar structures from a large structure database. Using the Elagolix structure as a template, six similar compounds were identified from EDULISS using UFSRAT and assayed for activity. From these six compounds, four were then used in a dose response curve to measure competitive binding to the GnRH receptor and were found to compete off 30-55% of the tracer peptide at 250 μM. Furthermore compound 15 was able to compete off 45% of the tracer at 50 μM. These compounds also demonstrated antagonist activity in a GPCR activation assay. This is a really encouraging result for our β_2 adrenergic based GnRH-R model structure, as stated; the only way to test model accuracy is by performing assays using suggested ligands.

Since we were able to identify four compounds all of which showed binding and antagonist activity then we are able to validate the model accuracy as well as the approach of using homology modelling for GPCR proteins.

Additionally quinolone compounds provided another compound similarity search approach. Again the compounds which were docked went on to show competitive activity when tested by assay. Of these compounds G10 was particularly notable,

with 34% of the tracer peptide displaced at a concentration of 33 μM. Ideally these compounds would then be tested using a dose response IP3 assay to assess the actual receptor activity inhibition.

However the latter part of this chapter does demonstrate one of the issues that can occur when using low sequence similarity structures as a modelling template. Namely that the template structure can introduce ligand bias. It was found that ligands to the β-adrenergic receptors scored more favourably in models created using the β_2-adrenergic receptor as a template than a model created from rhodopsin. The scoring was also lower than the GnRH receptor specific ligands which acted as a positive control.

All compounds identified could have their chemistry further optimised to increase binding activity as part of a QSAR analysis. From this novel scaffolds may be found which could make promising drug like molecules.

5.8 References

1. Chen, C., et al., *Discovery of Sodium R-(+)-4-{2-[5-(2-Fluoro-3-methoxyphenyl)-3-(2-fluoro-6-[trifluoromethyl]benzyl)-4-methyl-2,6-dioxo-3,6-dihydro-2H-pyrimidin-1-yl]-1-phenylethylamino}butyrate (Elagolix), a Potent and Orally Available Nonpeptide Antagonist of the Human Gonadotropin-Releasing Hormone Receptor.* Journal of Medicinal Chemistry, 2008. **51**(23): p. 7478-7485.

2. Struthers, R.S., et al., *Suppression of Gonadotropins and Estradiol in Premenopausal Women by Oral Administration of the Nonpeptide Gonadotropin-Releasing Hormone Antagonist Elagolix*, 2009. p. 545-551.

3. Cho, N., et al., *Discovery of a Novel, Potent, and Orally Active Nonpeptide Antagonist of the Human Luteinizing Hormone-Releasing Hormone (LHRH) Receptor.* Journal of Medicinal Chemistry, 1998. **41**(22): p. 4190-4195.

4. Kellenberger, E., et al., *Comparative evaluation of eight docking tools for docking and virtual screening accuracy*. Proteins: Structure, Function, and Bioinformatics, 2004. **57**(2): p. 225-242.

5. Betz, S.F., et al., *Determination of the Binding Mode of Thienopyrimidinedione Antagonists to the Human Gonadotropin Releasing Hormone Receptor Using Structure-Activity Relationships, Site-Directed Mutagenesis, and Homology Modeling.* Journal of Medicinal Chemistry, 2006. **49**(21): p. 6170-6176.

6. DeVita, R.J., et al., *Investigation of the 4-O-alkylamine substituent of non-peptide quinolone GnRH receptor antagonists.* Bioorganic & Medicinal Chemistry Letters, 1999. **9**(17): p. 2621-2624.

7. DeVita, R.J., et al., *Identification and initial structure-activity relationships of a novel non-peptide quinolone GnRH receptor antagonist.* Bioorganic & Medicinal Chemistry Letters, 1999. **9**(17): p. 2615-2620.

8. DeVita, R.J., et al., *Identification of neutral 4-O-alkyl quinolone nonpeptide GnRH receptor antagonists.* Bioorganic & Medicinal Chemistry Letters, 2004. **14**(22): p. 5599-5603.

9. Walsh, T.F., et al., *Potent antagonists of gonadotropin releasing hormone receptors derived from quinolone-6-carboxamides.* Bioorganic & Medicinal Chemistry Letters, 2000. **10**(5): p. 443-447.

Chapter 6. Identifying small molecule inhibitors using a LIDAEUS database mining approach

In Chapter 5 the use of similarity searching as a means of discovering novel compounds with affinity for the GnRH receptor was investigated and compounds were identified with antagonist activity. Whilst this has the advantage of similar compounds having similar chemical properties, the potential of finding new scaffolds for future drug development is reduced[1]. In this chapter an alternative approach of database mining was focussed on; a high throughput screening study was employed to dock and score over five million compounds against the GnRH receptor which generated compounds to assay for activity.

In this section of virtual screening three models of GnRH-R were used; two models were based on rhodopsin in the active and inactive states (as described in Chapter 3.5) and the last model was based on the β_2-adrenergic receptor in the inactive state (as described in Chapter 3.6). The changes in the active site conformation between all three models are highlighted in Figure 6-1. The site points in all models incorporated these residues into the docking area.

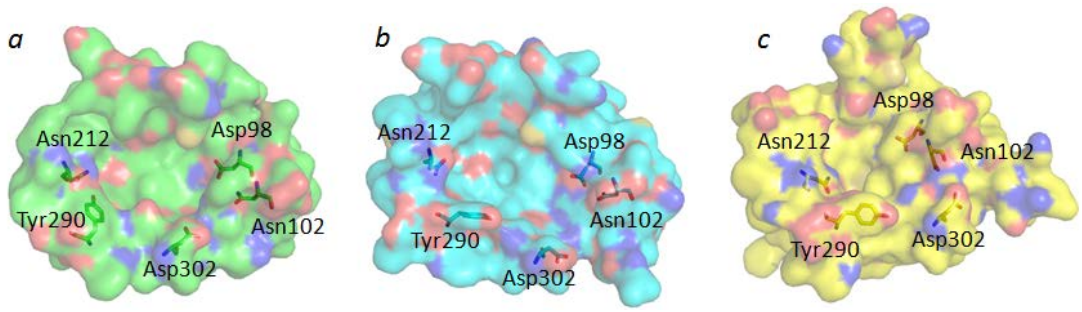


Figure 6-1. Positions of the key GnRH binding residues in each of the three model structures. (*a*) Rhodopsin derived GnRH-R in the active conformation. (*b*) Rhodopsin derived GnRH-R in the inactive conformation. (*c*) β_2-adrenergic receptor derived model in the inactive conformation.

6.1.1 Docking the GnRH-R model derived from rhodopsin in the active conformation with EDULISS compounds using LIDAEUS

In the active conformation GnRH-R is in the ligand bound state. This docking experiment allowed an attempt to dock small molecules into the receptor as a means of selecting for antagonists that have high affinity for the ligand bound conformation. This allows ligand selection based on displacement of the GnRH-I peptide.

Figure 6-2 shows the structure of the model which was used in these docking experiments as well as the positioning of the site points (site point generation is described in Chapter 2.4) which were used to guide the ligand positions to the GnRH-I active site.

The top 100 compounds from the virtual screening study were retained for further filtering.

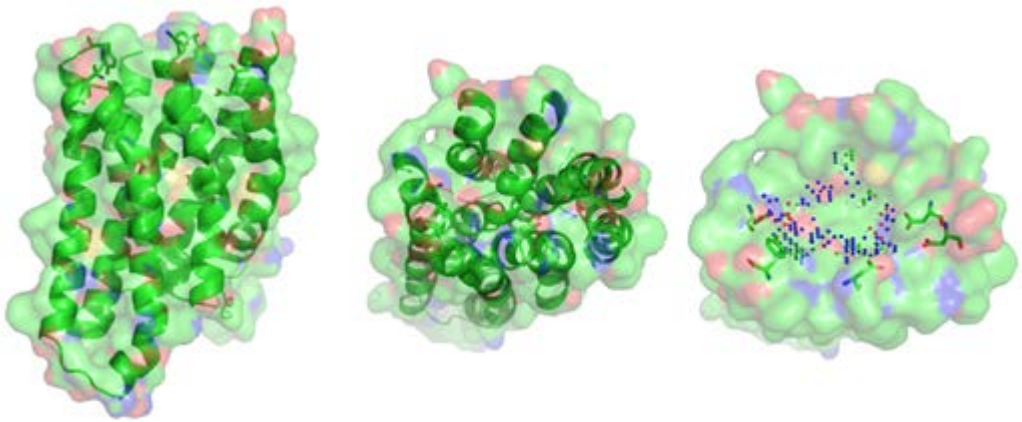


Figure 6-2. Structure of the active conformation rhodopsin derived GnRH receptor model used in this part of the docking study. Image shows side and top down views of protein chains with transparent surface as well as site point locations in comparison to key GnRH peptide binding residues.

6.1.2 Compound selection for assay

As the compounds from this screen were selected based on docking score rather than similarity to a reference structure compound selection was slightly different to the method used in Chapter 5. As discussed in Chapter 1.3.4, when a large chemical space is being investigated it is more efficient to cluster compounds by similarity, then to pick a compound to be a representative of each cluster rather than testing all members of the cluster. The method used to do this was to firstly increase the drug likeness of the candidate molecules by applying the Lipinksi rule of five to the top 100 compounds; this reduced the compound number from 100 to 63. The remaining 63 compounds were then clustered (as described in Chapter 2) according to Soergel distance which is the compliment of the Tanimoto co-efficient[2] (most similar compounds score closest to 0) using Daylight fingerprints (Figure 6-3).

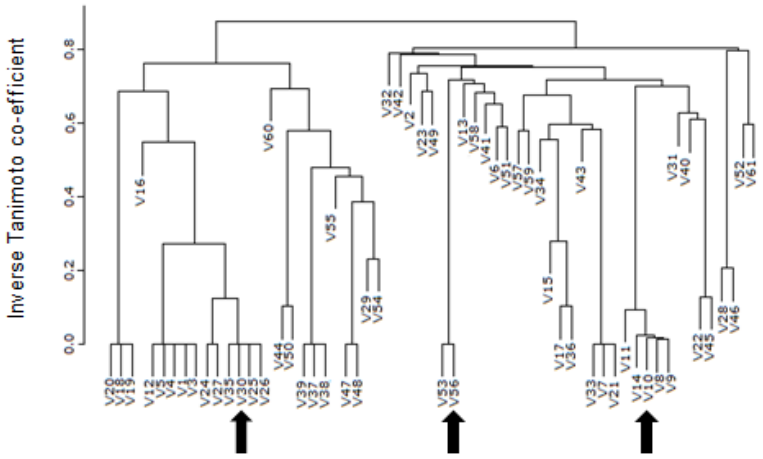
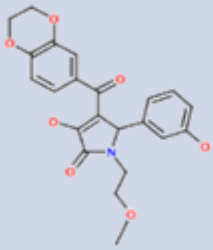
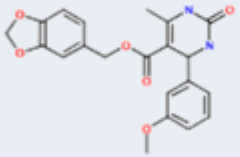
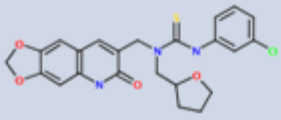


Figure 6-3. Clustering of active state GnRH-R from homology modelling with rhodopsin. Picking compounds from a clade as a representative of biological activity from within the group is an efficient way of rapidly screening potential inhibitors. Black arrows indicate positions of compounds selected.

Upon completion of the clustering stage three compounds (Table 6-1) were selected from the remaining selection to test. The compounds were selected based on criteria such as predicted solubility, availability and LIDAEUS score.

Table 6-1. Compounds selected for assay from the active state structure search for antagonist ligands.

Compound letter	Structure	Mw (Da)	Manufacturer ID	LIDAEUS score
A		423.462	STOCK4S-27724	-50.8
B		396.393	5615633	-49.4
C		471.956	ASN 03776013	-53.3

6.1.3 Assay of potential antagonist compounds from active state rhodopsin homology model

The assays of the compounds from the screening described in this chapter are the same as those used previously. Compounds were assayed at a concentration of 100µM in a single point inositol (1,4,5) trisphosphate (IP3) assay as described in

Chapter 2.10.4 and Chapter 5.5. The assay is performed in the presence of 5% DMSO with a GnRH-I stimulated negative control included to allow visualisation of antagonist/agonist activity, the results are displayed in Figure 6-4.

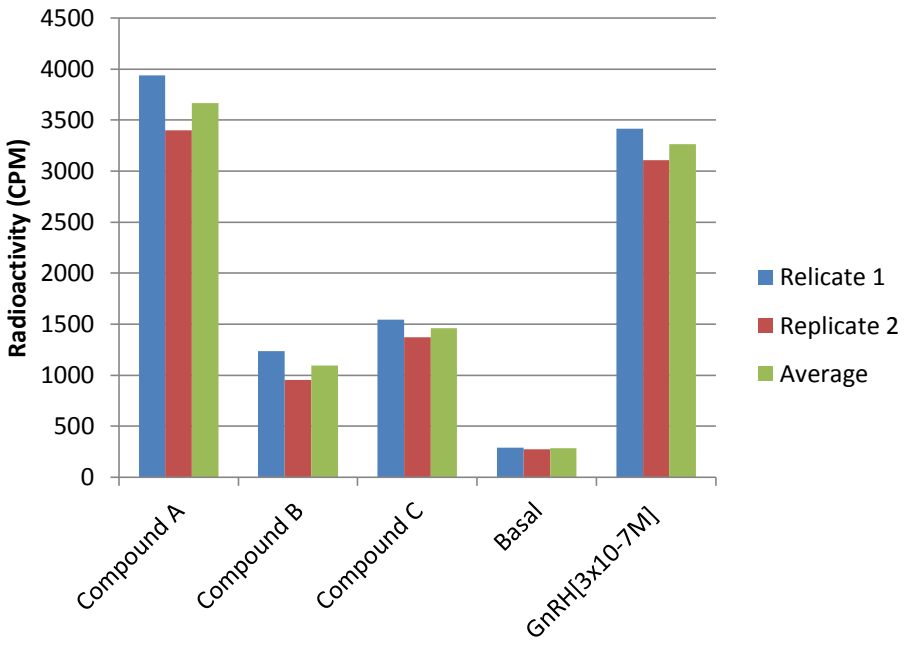


Figure 6-4. Results of IP assay for 100 µM antagonist compounds from rhodopsin active state model. Each condition was set up in duplicate and the average radioactivity reading was calculated. GnRH I was used as a positive control.

Of the three compounds tested only compound A showed no significant IP inhibition, however this compound did appear to have small agonistic properties. Compounds B and C both showed an inhibition of IP production that was greater than 50% at the 100 µM concentration. These compounds were tested in a radioligand binding assay to assess how well they could compete GnRH-I from the active site (Figure 6-5). The protocol for this is described in Chapters 2.10.3 and 5.5.1.

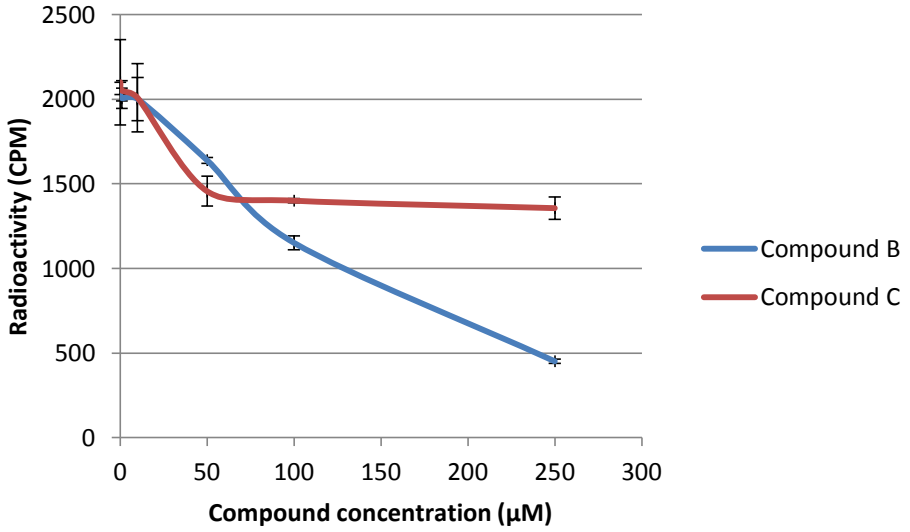


Figure 6-5. Radioligand binding assay of compounds produced from EDULISS screening of the GnRH-R rhodopsin model.

The binding assay results showed that compound B was able to successfully compete with the radiolabelled peptide with increasing concentration. Compound C was less active, 25% of the radiolabelled peptide was competed off at 50 µM, however with increasing concentration beyond this point no further effect was observed. This coincides with the observation that the compound became more insoluble as final compound concentration increased. Compounds which form aggregates are an established source of false positives in small molecule screening[3, 4], therefore care must be taken when such behaviour is observed. The non-specific inhibition which can occur as a result of aggregate formation is thought to occur by sequestering enzyme molecules within the aggregation which reduces the ability to bind the activating factor[5, 6]. This cannot take place in this assay because the receptor is anchored into the cell membrane.

Virtual screening using the rhodopsin derived active state model allowed the discovery of two compounds which were able to demonstrate inhibition of IP3 production and thus receptor inhibition as well as competitive inhibition with a high affinity tracer peptide. A weak agonist was also found which suggests that structures which are far smaller than the native peptide are still able to form key interactions capable of activating the receptor.

6.2 Docking the GnRH-R model derived from rhodopsin in the inactive conformation with EDULISS compounds using LIDAEUS

The conformation of the rhodopsin derived inactive state GnRH-R model (Figure 6-6) was designed to mimic the active site when in a ligand-free state, this theoretically allows small molecules to be found which can bind to this state and prevent GnRH-I peptide binding. As the ligand molecules in EDULISS are too small to bind to all of the residues involved in agonist binding, activation of receptor signalling should be avoided whilst still restricting GnRH I access to the binding site. The construction of this model is detailed in Chapter 3.5.

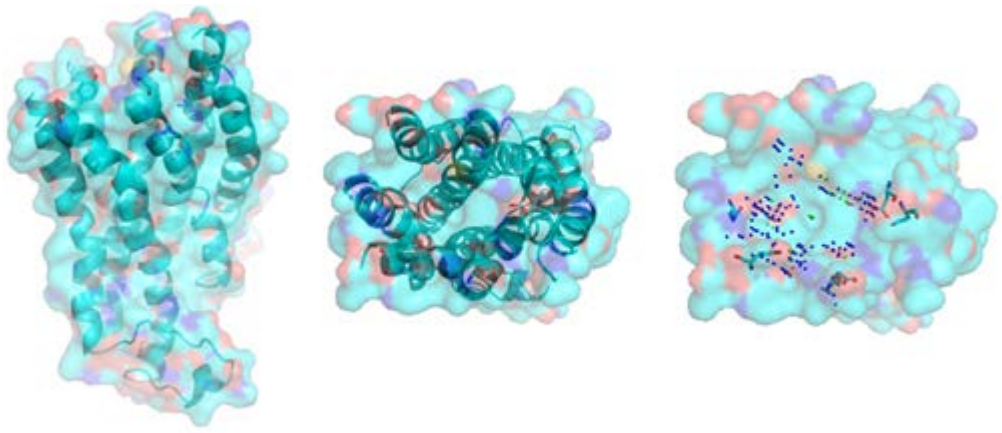


Figure 6-6. Model structure used in this docking experiment which represents GnRH-R in the inactive state based on the rhodopsin structure. This model differs from the active site rhodopsin derived model by 1.7Å in a RMSD fit of 97 Cα atoms. Rotation of residues 290 and 302 are a feature of this model.

The docking process that was followed is the same as that used for the active state model, however the site points differed slightly in this case. As the target is the peptide free structure the site points do not need to be clustered so heavily towards the active site residues; providing the small molecule can make a strong interaction with any residue located close to the active site then it should be sufficient to block entry of the peptide and prevent receptor activation. After docking was complete the top 100 hits were retained for further filtering.

6.2.1 Compound selection for assay

Of the top 100 compounds, 43 passed the Lipinksi rule of five. These were clustered using the same method that was applied to the hits obtained for the active state rhodopsin derived GnRH-R model and the results are shown in Figure 6-7.

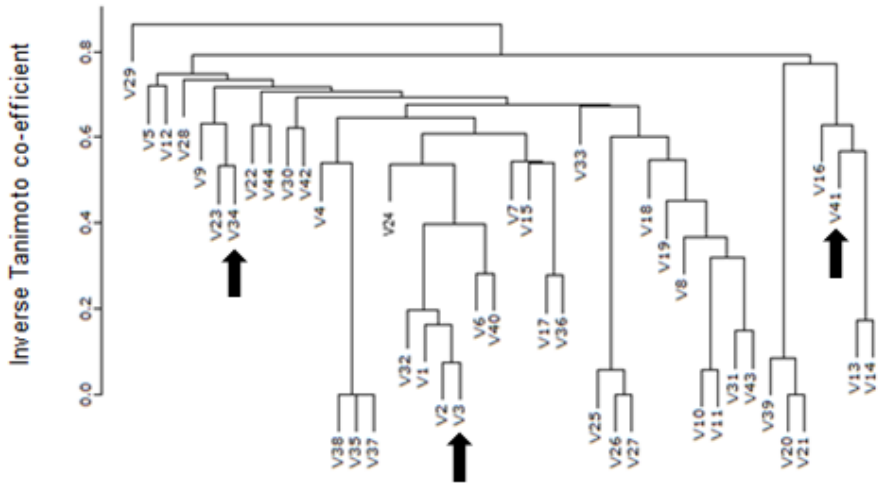


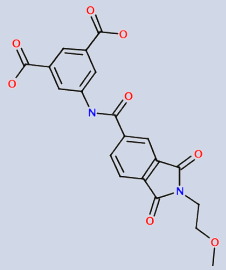
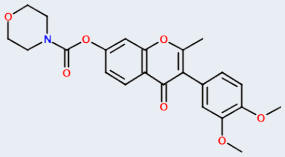
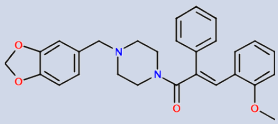
Figure 6-7. Clustering of inactive state GnRH-R from homology modelling with rhodopsin. Picking compounds from a clade as a representative of biological activity from within the group is an efficient way of rapidly screening potential inhibitors. Black arrows indicate the position of the compounds selected.

From the 43 compounds which passed the Lipinski criteria three compounds were selected to assay. The structures of these compounds are shown in Table 6-2 overleaf. LIDAEUS scores from this selection are lower than those obtained for the active state model, however this was because the highest scoring compounds were not in stock with the suppliers.

6.2.2 Assay of potential antagonist compounds from inactive rhodopsin homology model

The assay method for the three chosen compounds for the inactive state rhodopsin derived model was the same as that used for the active state model; compounds were tested using IP3 inhibition and competition binding against a radiolabelled peptide. A single concentration IP3 assay was used first to quickly determine the potency of antagonism of the selected compounds (Figure 6-8).

Table 6-2. Compounds selected for assay from the ground state structure search for inverse agonists.

Compound letter	Structure	Mw (Da)	Manufacturer ID	LIDAEUS score
D		412.350	7961791	-42.541
E		425.431	STOCK4S-46535	-42.162
F		457.541	6R-0209	-41.908

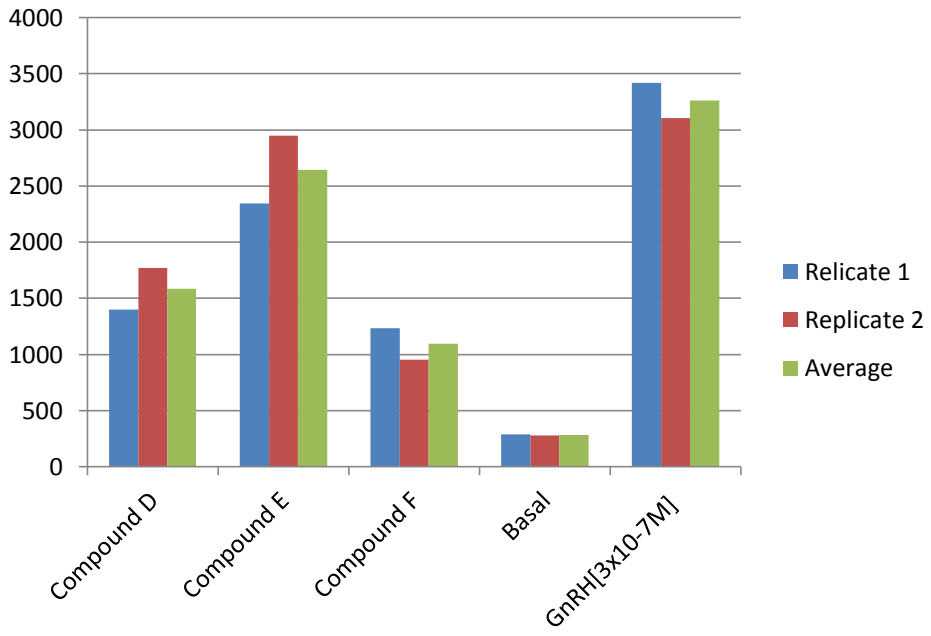


Figure 6-8. IP inhibition results for antagonist compounds derived from the inactive GnRH-R rhodopsin homology model.

The one point assay results (Figure 6-8) showed that all of the selected compounds had antagonist properties. Compound E reduced IP production the least with a reduction of 20% at 100 µM concentration, compounds D and F were slightly more potent causing a decrease of 53% and 65% respectively.

As compounds D and F showed antagonist activity in the single concentration IP assay they were then tested in a dose-response binding assay with a compound concentration range of 0.01-250 µM (Figure 6-9).

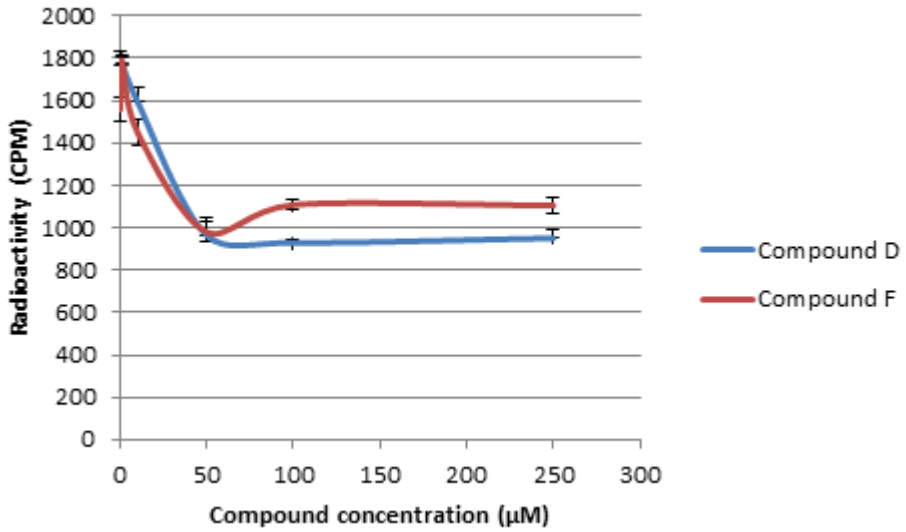


Figure 6-9. Binding assay of compounds derived from rhodopsin homology model.

The binding assay indicates that both compounds D and F are able to compete the tracer peptide from the receptor binding site. This supports the view that the antagonist effects which were observed in the IP3 assay occur because of specific binding to the receptor rather than a non-specific decrease in IP3 production which could have been attributed to cell toxicity[7].

No agonists were found from this round of screening. Whilst it would be attractive to attribute this as a factor of the difference between the active and inactive models it is also possible that the 41 compounds which were not assayed could have agonist properties that were not possible to observe.

6.3 Docking the GnRH-R model derived from the β_2 adrenergic receptor in the inactive conformation with EDULISS compounds using LIDAEUS

This model structure (Figure 6-10) is similar to the model that was discussed in section 6.2 as it is also an inactive state model, however it differs in the template model which was used. Using the β_2 adrenergic receptor as the template structure allows the transmembrane helixes to be modelled in a way which accommodates the association and dissociation of a diffusible ligand[8]. This should create a model which is more representative of the GnRH-R structure although it does share the same limitations as previous crystal structures; the crystal structure can only capture a single "moment" of the receptor structure oscillation[9] which was discussed in Chapter 1.

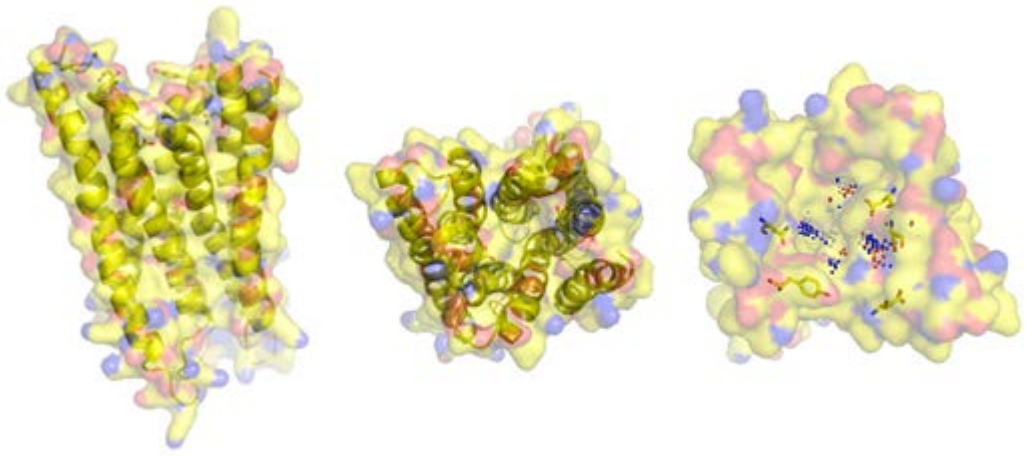


Figure 6-10. Structure of the inactive GnRH receptor model derived from β_2AR. This model has an RMSD value of 3.7Å when aligned against the inactive GnRH-R model derived from rhodopsin.

The creation of site points and the subsequent LIDAEUS screening were performed as described for the previous two models, in this model the site points tended to cluster in a more centralised location over the previous models. The top 100 compounds were selected for further filtering.

6.3.1 Compound selection for assay

The top 100 compounds were filtered by the rule of five which left 74 compounds remaining to be clustered according to Soergel distance (Figure 6-11).

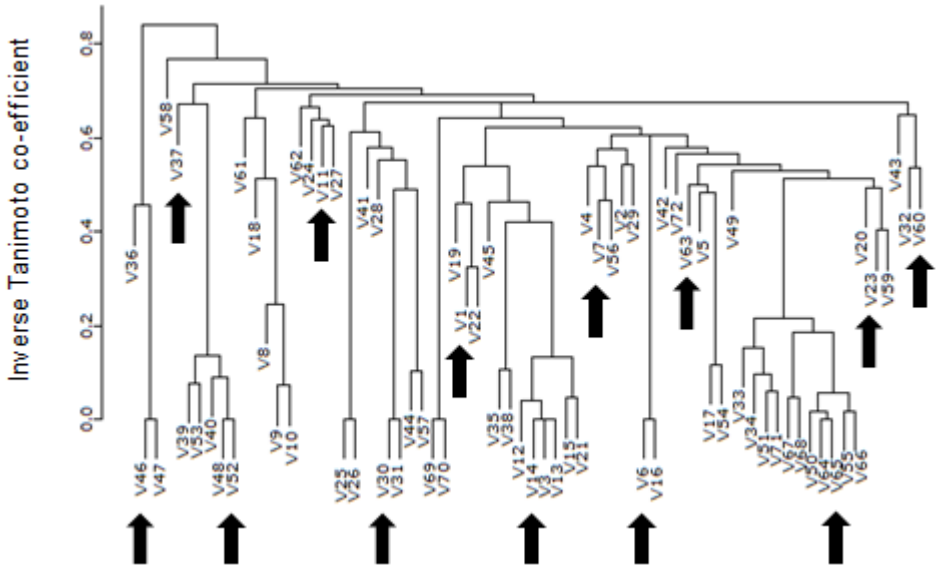
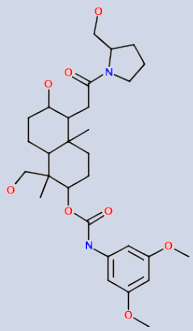
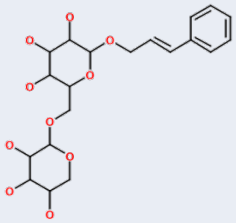
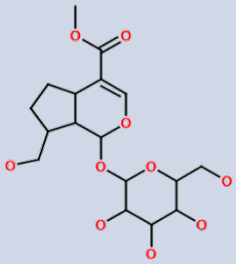
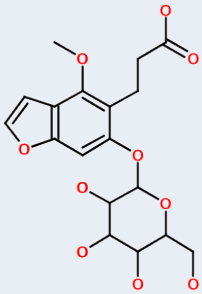
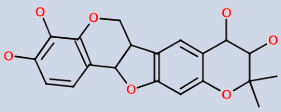
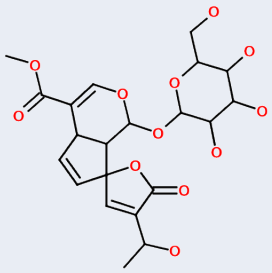
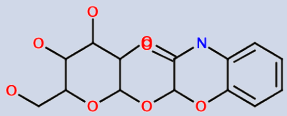
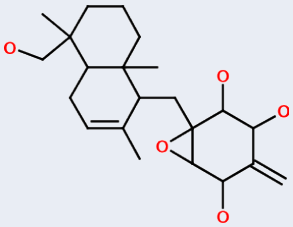
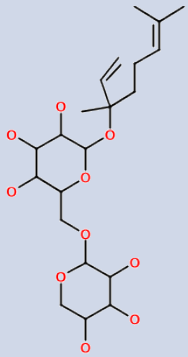
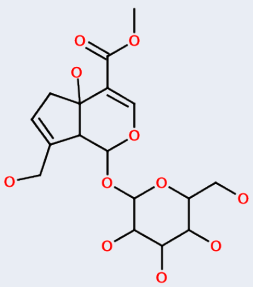
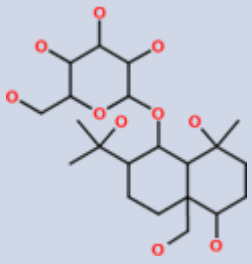


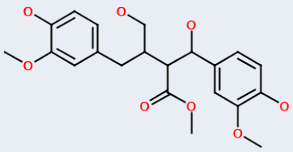
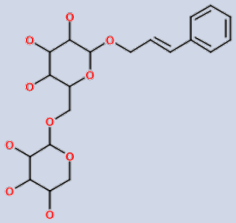
Figure 6-11. Clustering of inactive state GnRH-R from homology modelling with β_2AR. Picking compounds from a clade as a representative of biological activity from within the group is an efficient way of rapidly screening potential inhibitors. Black arrows indicate compounds selected.

From the 74 compounds which passed the Lipinski criteria thirteen compounds were selected to assay, these are shown in Table 6-3. The LIDAEUS scores from this selection were more favourable than those obtained for the compounds from the equivalent rhodopsin structure.

Table 6-3. Compounds selected for assay from LIDAEUS β_2AR GnRH-R model

No	Structure	Mw (Da)	Manufacturer ID	LIDAEUS Score
1		548.66	NAT8-282963	-42.804
2		428.43	NP-001299	-56.132
3		390.382	NP-002450	-48.691
4		398.361	NP-003411	-48.164
5		372.36	NP-005953	-52.2069

6		470.42	NP-013761	-43.6059
7		327.28	NP-007338	-42.7076
8		378.5	NP-004042	-46.4451
9		448.5	NP-002635	-58.4059
10		404.366	NP-007310	-47.1548
11		450.52	NP-002436	-46.334

12		406.426	NP-002445	-42.9256
13		428.43	NP-015267	-56.132

The compounds contained a number of hydroxyl groups as well as one compound with an epoxide ring. Toxicophore screens would usually remove epoxide ring containing compounds due to their high reactivity with proteins[10], although a number of very successful drugs such as Clopidogrel (for treatment of cardiovascular conditions) and Omeprazole (for treatment of gastric disorders) contain toxicophores[11].

6.3.2 Assay of potential antagonist compounds from β_2 adrenergic receptor homology model

As in the previous assays the compound concentration of the single point IP assay was increased to 100 µM for testing. Of the 13 compounds from the β_2AR model there were three compounds which demonstrated antagonist behaviour in the IP assay. These were compounds 5, 8 and 12 (Figure 6-12).

Ideally these compounds would have been tested in a dose-response assay as previously used for the rhodopsin derived models, however this was not possible in this case due to time constraints.

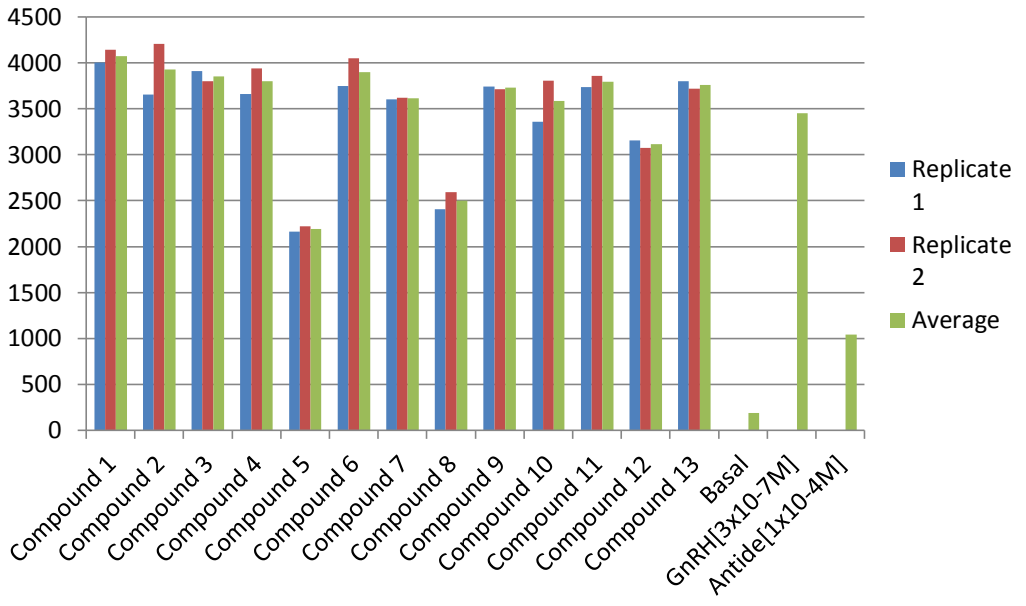


Figure 6-12. One point assay results of compounds selected using the β_2AR model structure.

6.4 Comparison of results from each model structure

Taking the most active compound from each of the three model structures and using AutoDock the possible interactions for each compound were mapped within their respective model structure (Figure 6-13).

The affinity of the best poses of each compound and their interactions with the receptor model are shown in Table 6-4 and highlights a few points for discussion. Whilst compound D was the lowest affinity ligand by AutoDock score, it performed the most favourably at lower compound concentration in the assays. The compound has the possibility of forming twice the number of hydrogen bonds of the other two compounds and this may be a key factor in the performance of compound D.

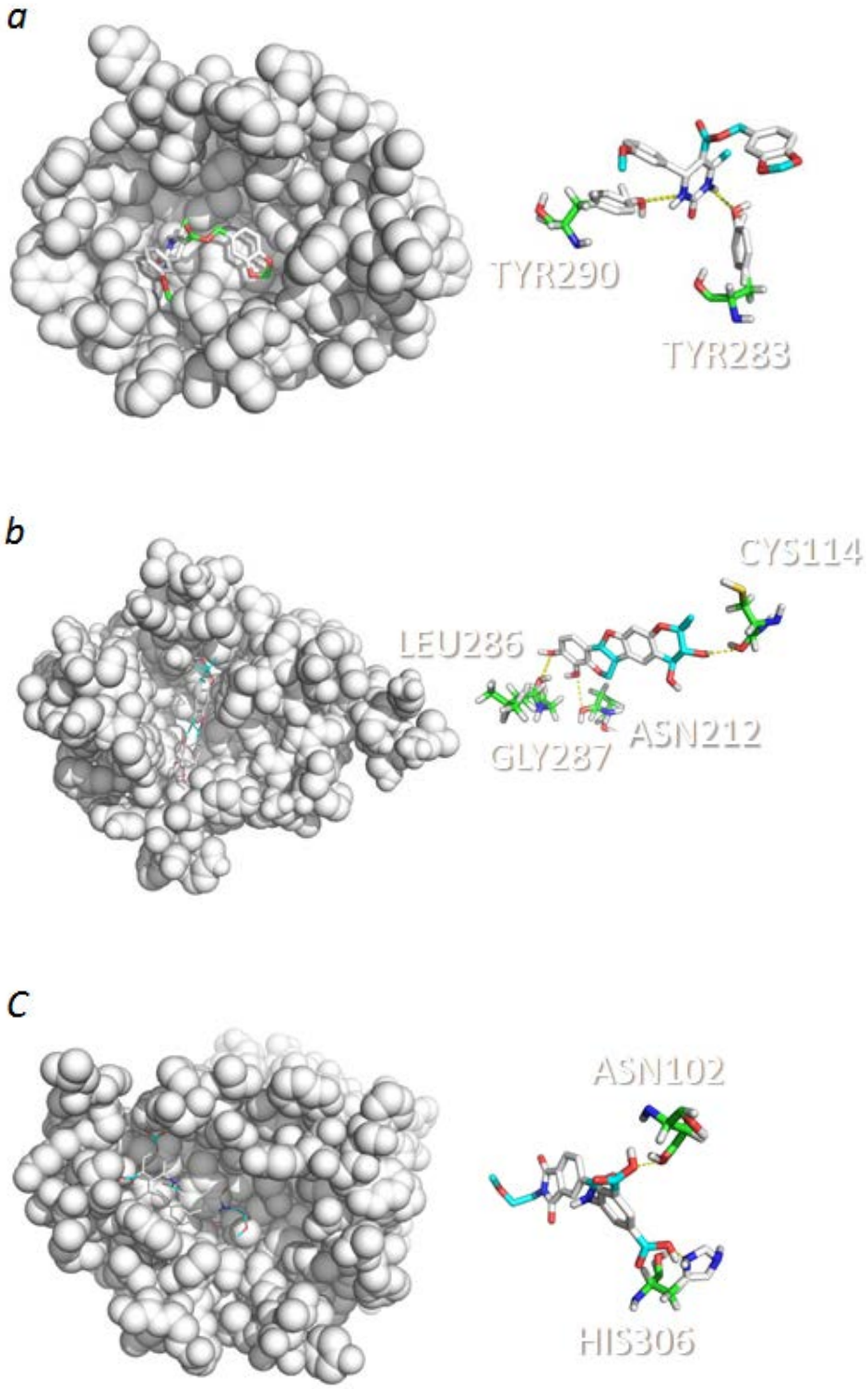


Figure 6-13. Comparison of docked positions of the most active compounds. (a) Rhodopsin model: active conformation with compound B. (b) Rhodopsin model: inactive conformation with compound D. (c) β_2AR model: inactive conformation with compound 5.

Table 6-4. Affinity scoring for each model compound which produced the most inhibition.

Compound	Docking affinity (kcal/mol)	Hydrogen bonds
B	-8.3	2
D	-7.5	4
5	-8.1	2

An interesting outcome of this study is the discovery of eight compound structures which showed inhibition of IP3 response from the GnRH receptor which were found using three different model structures. From these eight compounds two showed antagonist activity in two different assays which quantifies them as not being false positives resulting from an artifact in the assay.

The variation in model structure and the relative indifference in small molecule affinities is perhaps a reflection on the current state of the majority of GPCR structural information. A fuzzy picture with defined boundaries, just enough to provide a subtle clue but lacking the brush strokes to provide true clarity.

6.5 References

1. Eckert, H. and Bajorath, J., *Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches.* Drug Discovery Today, 2007. **12**(5–6): p. 225-233.

2. Fechner, U. and Schneider, G., *Evaluation of Distance Metrics for Ligand-Based Similarity Searching.* ChemBioChem, 2004. **5**(4): p. 538-540.

3. Roche, O., Schneider, P., Zuegge, J., Guba, W., Kansy, M., Alanine, A., Bleicher, K., Danel, F., Gutknecht, E.-M., Rogers-Evans, M., Neidhart, W., Stalder, H., Dillon, M., Sjögren, E., Fotouhi, N., Gillespie, P., Goodnow, R., Harris, W., Jones, P., Taniguchi, M., Tsujii, S., von der Saal, W., Zimmermann, G., and Schneider, G., *Development of a Virtual Screening Method for Identification of "Frequent Hitters" in Compound Libraries.* Journal of Medicinal Chemistry, 2001. **45**(1): p. 137-142.

4. Keserű, G.M. and Makara, G.M., *Hit discovery and hit-to-lead approaches.* Drug Discovery Today, 2006. **11**(15–16): p. 741-748.

5. Feng, B.Y., Simeonov, A., Jadhav, A., Babaoglu, K., Inglese, J., Shoichet, B.K., and Austin, C.P., *A High-Throughput Screen for Aggregation-Based Inhibition in a Large Compound Library.* J. Med. Chem., 2007. **50**(10): p. 2385-2390.

6. Feng, B.Y., Toyama, B.H., Wille, H., Colby, D.W., Collins, S.R., May, B.C.H., Prusiner, S.B., Weissman, J., and Shoichet, B.K., *Small-molecule aggregates inhibit amyloid polymerization.* Nat Chem Biol, 2008. **4**(3): p. 197-199.

7. Cummings, M.D., Arnoult, É., Buyck, C., Tresadern, G., Vos, A.M., and Wegner, J.K., *Preparing and Filtering Compound Databases for Virtual and Experimental Screening*, in *Virtual Screening*2011, Wiley-VCH Verlag GmbH & Co. KGaA. p. 35-59.

8. Audet, M. and Bouvier, M., *Insights into signaling from the β2-adrenergic receptor structure.* Nat Chem Biol, 2008. **4**(7): p. 397-403.

9. Millar, R.P. and Newton, C.L., *The Year In G Protein-Coupled Receptor Research.* Molecular Endocrinology, 2010. **24**(1): p. 261-274.

10. Lemke, T.L., *Review of organic functional groups: introduction to medicinal organic chemistry*2011: Lippincott Williams and Wilkins.

11. Xu, J.J. and Urban, L., *Predictive Toxicology in Drug Safety*2010: Cambridge University Press. 113-114.

Chapter 7. Biochemical analysis and structural characterisation of Post Synaptic Density protein 95 (PSD95)

7.1 Introduction

As described in Chapter 1 the goal of targeting the SH3 domain of PSD95 is to inhibit the scaffold activity that is linked to the transmission of neuropathic pain after sciatic nerve injury[1, 2]. This is a significant area of research as current analgesic treatment of sciatica often involves the use of strong opiates which are habit forming after extended use[3].

The pathway involving PSD95 controlled inflammatory response involves the recruitment of phosphatidylinositol-3-kinase-C2α (PI(3)K-C2α) which is a signalling molecule belonging to the PI(3)K family[2]. This signalling molecule contains a proline rich sequence FPLWKLPGFPNRMVLG at residues 1477-1492, which can associate with PSD95 (Figure 7-1) in the spinal cord increasing inflammatory sensitisation[2]. The aim of the work described in this chapter is to identify small molecules capable of inhibiting the interaction of PSD95 and PI(3)K-C2α to act as strong pain killers.

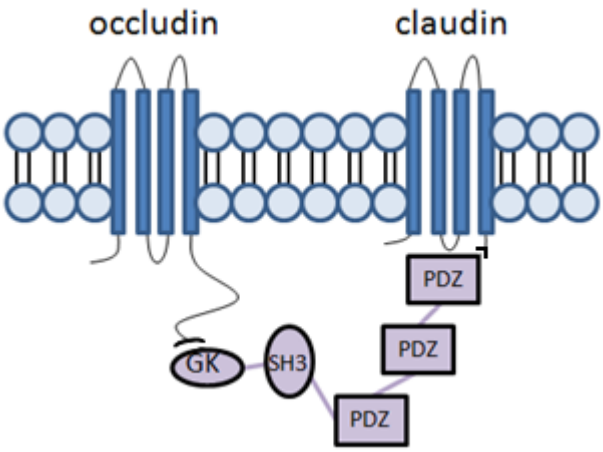


Figure 7-1. A proline rich region of PI(3)KC2α is able to associate to the SH3 domain of PSD95 and control inflammatory response.

7.1.1 Canonical SH3 domains

Src-homology 3 (SH3) domains are found in a number of signalling pathways and aid signal transduction through protein-protein interactions[4]. SH3 domains are found in a number of eukaryotes, it has been found that in the human genome there are over 500 SH3 domains[5]; this makes it an attractive target for development of ligand molecules using virtual screening.

The SH3 domain is comprised of around 60 residues and can be split into two classes, the first class binds proline rich motifs of sequence pΦPxP while the second class binds the motif ψPpΦPΦ where P represents a proline residue, p represents a proline residue which is not highly conserved, Φ represents a hydrophobic residue and ψ represents an aliphatic residue[6].

The binding pocket of the SH3 domain consists of a small area (~400Å^2) between the n-Src and RT loops (Figure 7-2), peptide association is weak (generally not more

than 50 μM[7]) and type I or II peptide specificity is determined by the positioning of a well conserved tryptophan in the specificity pocket[8].

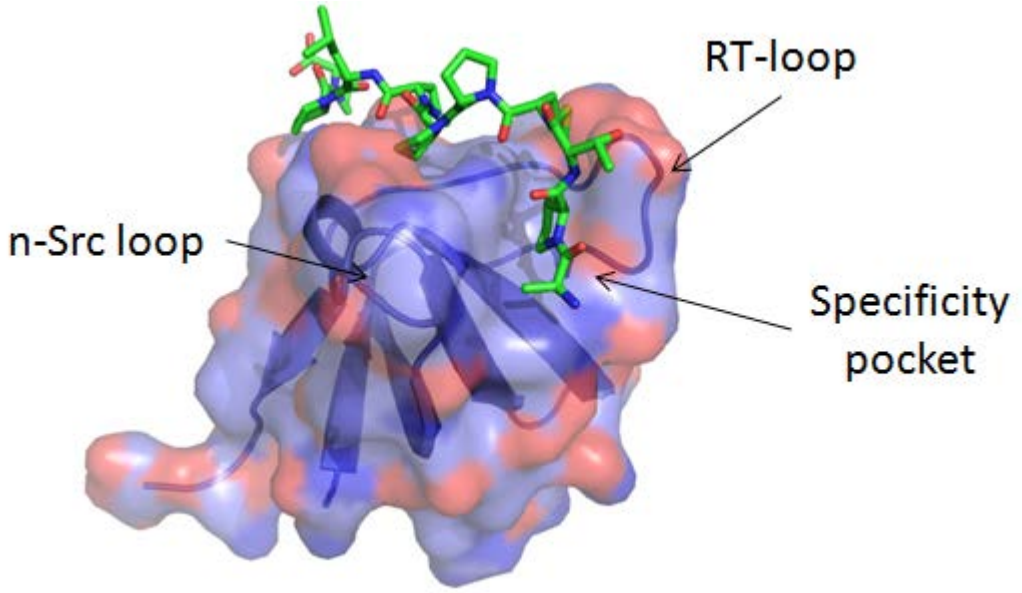


Figure 7-2. SH3 domain from ABL tyrosine kinase (PDB ID 1ABO). The proline rich peptide binds in a class I conformation in a shallow pocket that is created between the n-Src loop and the RT-loop. The specificity pocket contains an integral tryptophan which determines peptide class binding.

Binding of proline rich sequences to the SH3 domain causes discrete conformational change within the protein. This was observed in NMR analysis of the SH3 domain of SEM5 where chemical shifts were observed for residues localised in the RT and n-Src loops. These residues were F163, D164, F165, N166, Q168, E169, E172, N190, W191, P204, N206, Y207, S170, G171, L173, F175, W192 and S205 which are shown in Figure 7-3[9]

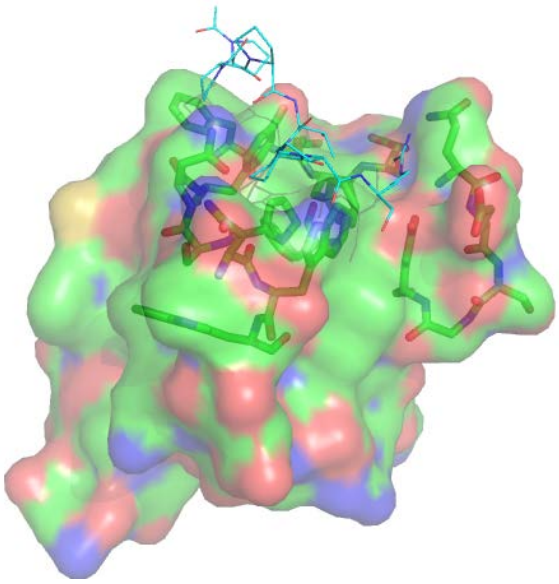


Figure 7-3. Surface view of SEM5 protein with residues which undergo chemical shift on ligand binding shown as sticks.

The ligand binding is coupled to reduced flexibility within the RT-loop in a cooperative mechanism; in the unbound state there is greater flexibility in the RT loop than any other region in the SH3 domain. Once the proline rich peptide begins to associate, the flexibility of the RT loop is greatly reduced and the ligand affinity increases.

Use of peptide sequences to disrupt SH3 domain interactions in treatment of conditions such as prostate cancer have been reported[10], however peptides are less attractive than small molecules as potential drugs. Small molecule inhibitors of SH3 domains have also been reported in the literature[11, 12], the small molecule reported (Figure 7-4) acted as a broad range SH3 interaction disruptor against a number of SH3 domain proteins.

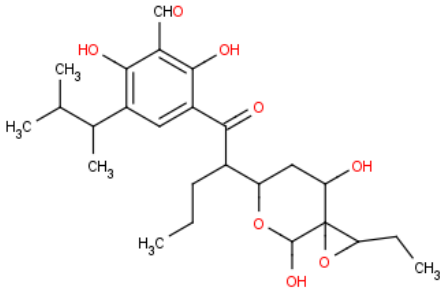


Figure 7-4. Structure of the SH3 domain small molecule inhibitor UCS15A[12].

Widespread disruption of SH3 domain interactions is undesirable, however the UCS15A molecule is a useful proof of concept that the SH3 domain is druggable and further modification of the scaffold may be a means of increasing specificity.

7.2 Identifying a druggable construct of PSD95

As discussed in Chapter 1, the SH3 domain of PSD95 has a unique intertwined structure involving the SH3 domain and from the Guanylate Kinase (GK) domain. A ligand bound crystal of the PSD95 SH3 has not yet been obtained, so initial computational study was performed to assess which areas of the structure would be most favourable to bind to.

Using the program STP (as described in Chapter 2) a map of ligand binding propensity was generated for the SH3-Hook-GK fragment of PSD95 as shown in Figure 7-5. The program highlights the trench that forms between the SH3 and GK domains as a likely binding area. The area that scores most favourably (red) is the area which incorporates a highly conserved ASP residue which when mutated in other SH3 domains eliminates PxxP binding[13].

Despite the apparent similarities between the SH3 domain of PSD95 and the canonical SH3 domain there are subtle differences between the structures as shown in Figure 7-6.

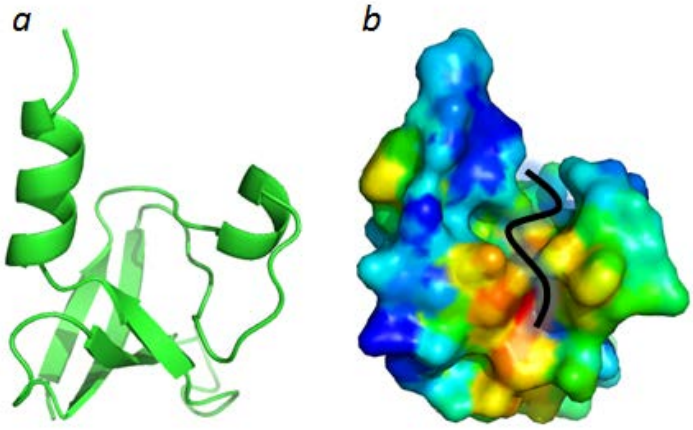


Figure 7-5 Binding propensity of the SH3-Hook-GK region of PSD95 as coloured by STP. The program colours areas which are most likely to act as a binding site as red whilst the areas least likely to act as a binding site as blue. (*a*) Ribbon diagram illustrates the structural arrangement of the protein fragment (*b*) Surface area view coloured to indicate possible binding regions. The black line indicates the typical PxxP motif binding orientation to traditional SH3 proteins.

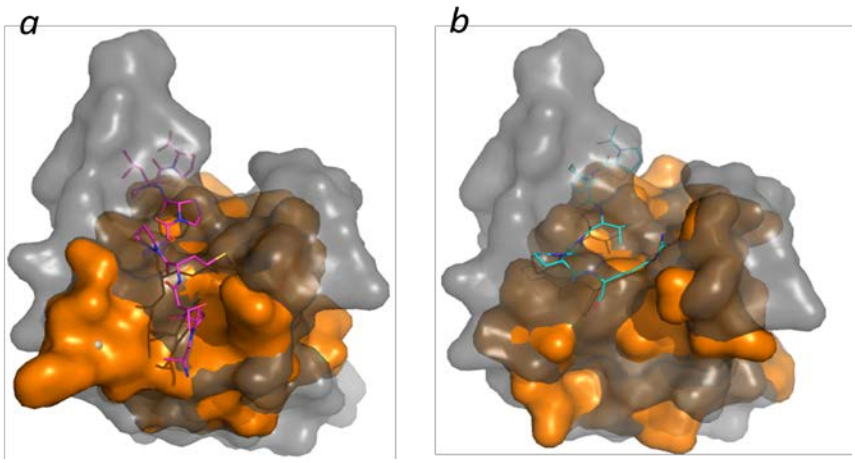


Figure 7-6. Comparison of binding orientations of class I and II SH3 domains to PSD-95. In both illustrations PSD-95 is shown as a translucent grey structure aligned with the SH3 domain it rests on and the canonical SH3 domain is shown in orange. RMSD values for structure alignments are 0.7Å and 0.6 Å for the class 1, 1ABO structure (a) and class 2, 1SEM structure (b) alignments respectively.

A comparison of the canonical class I and class II peptide binding SH3 domains with the PSD95 SH3 domain shows a non-canonical peptide binding interface (Figure 7-6). Overlay of the PSD95 SH3 with the SH3 domain of 1ABO (Figure 7-6a) shows occlution of the peptide binding interface by an alpha helix. This is also evident when PSD95 is overlaid with the class II 1SEM domain (Figure 7-6b).

This highlights a fundamental difference in binding orientation given the way that the PSD-95 molecule clashes with the ligand structure of both the class I and II domains. This suggests that whatever molecules were found to bind to PSD-95 may have some inherent specificity to MAGUKs, depending on the region targeted, which is very beneficial from a drug design aspect.

7.2.1 Alignment of the PSD95 SH3 domain with canonical SH3 domains

A pairwise alignment created using ClustalW of the PSD95 SH3 domain against 8 other class 1 and class 2 SH3 domains indicates higher than expected conservation between the sequences (Figure 7-7).

The percentage identity between the sequences in comparison to PSD95 SH3 ranges from 14% to 29% (Table 7-1). This range in identity score is surprising for such a highly conserved domain particularly given the low RMSD values for the 3D structural alignments. The highest RMSD value was 1.52Å against 2J7I which is an SH3 domain of the CMS (Cas ligand with multipleSrc homology 3 (SH3) domains) family.

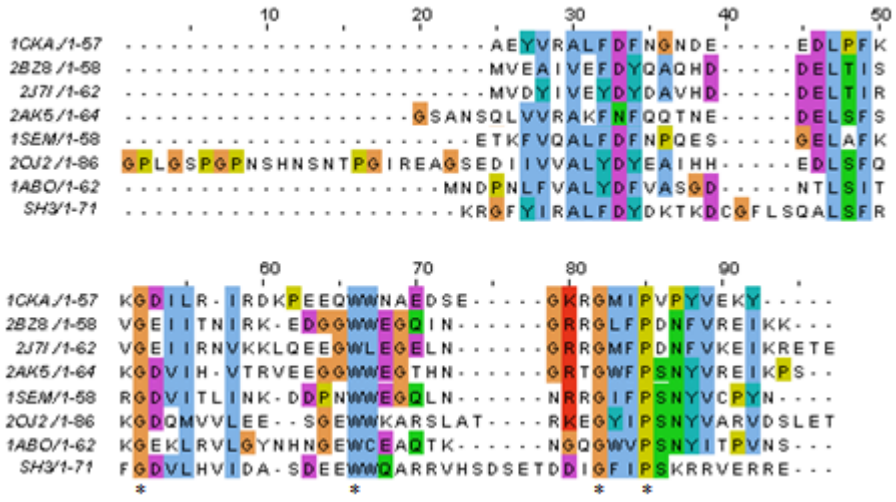


Figure 7-7. Sequence alignment of canonical SH3 domains (both Class I and II) and PSD-95 SH3 domain. The sequence corresponding to PSD95 SH3 is labelled as "SH3". Residues that are fully conserved across all sequences are highlighted by an asterisk.

Table 7-1. Identity scores between PSD95 SH3 and 7 other class 1 and class 2 SH3 domain proteins.

PSD95 SH3 aligned against	Identity (%)	RMSD (Å)	Cα atoms included in fit
1ABO	14	0.69	32
2BZ8	17	0.95	25
2J7I	19	1.52	23
2AK5	21	0.63	34
1SEM	24	0.57	32
1CKA	28	0.57	34
2OJ2	29	1.08	39

These proteins recognised specific proline/arginine sequences of P*XXX*PR[14], 2BZ8 also binds similar proline/arginine sequences[15]. The RMSD value of 1.08 for 2OJ2 is also a point of interest as this protein is able to bind to a non-typical proline rich sequence of xxxPpxP[16]. This highlights the diverse roles that the SH3 domain plays

in protein interactions; where a number of subtle changes from a core motif allow it to play an integral part in a number of scaffold functions.

To further investigate the diversity within the ligands of SH3 domains, 6 peptide structures which bind to SH3 domains from Table 7-1 were overlaid (Figure 7-8).

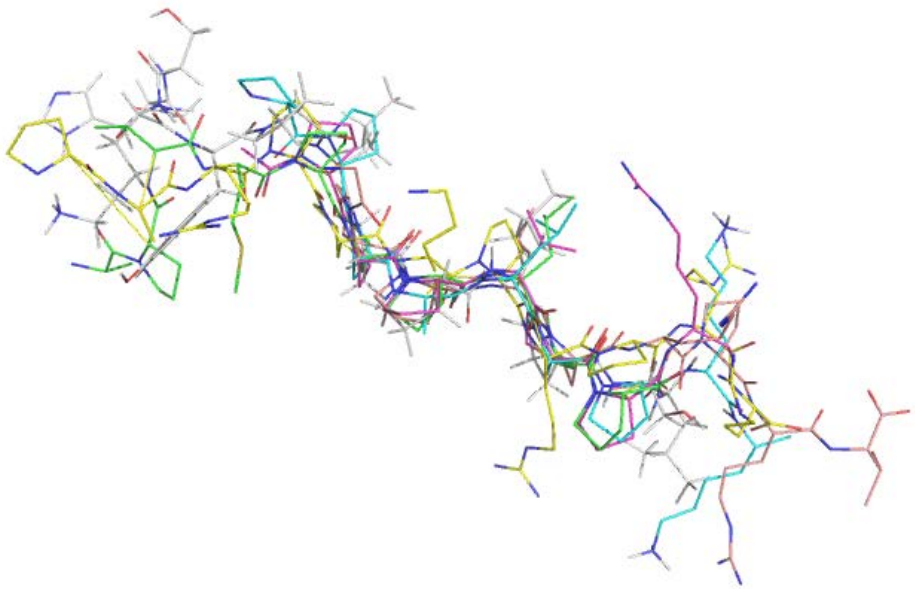


Figure 7-8. Overlay of peptides structures which bind to SH3 domains aligned in Figure 7-7. 1ABO (green), 1CKA (cyan), 1SEM (magenta), 2BZ8 (yellow), 2JZI (pink), 2OJ2 (white).

To further investigate the peptide sequences which have been shown experimentally to bind to SH3 domains an extensive literature survey was performed to identify peptide sequences which are able to bind SH3 domains[16-31]. From this survey a database of 214 SH3 peptides was compiled. The frequency of each of the amino acids (where possible three residues before and after the PxxP motif) was established (Figure 7-9). The nomenclature used is as follows; residues found N-terminal to the PxxP motif are denoted as -3, -2 and -1 whilst residues C-terminal to the PxxP motif are denoted as +1, +2 and +3.

The analysis was extended to include only the two “x” residues located between the prolines within the PxxP motif. The nomenclature used is as follows Px_1x_2P which is shown if Figure 7-10.

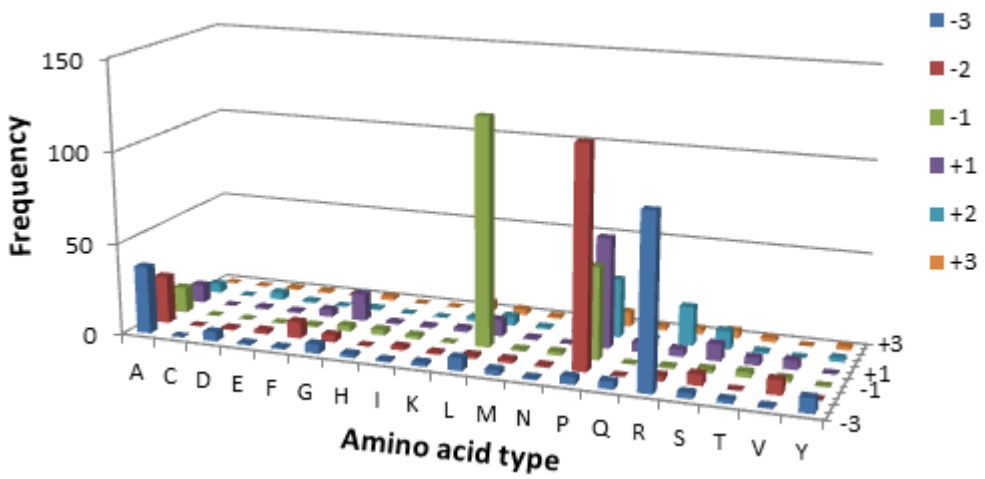


Figure 7-9. Amino acid distribution of the three residues before and after the PxxP motif binding peptides.

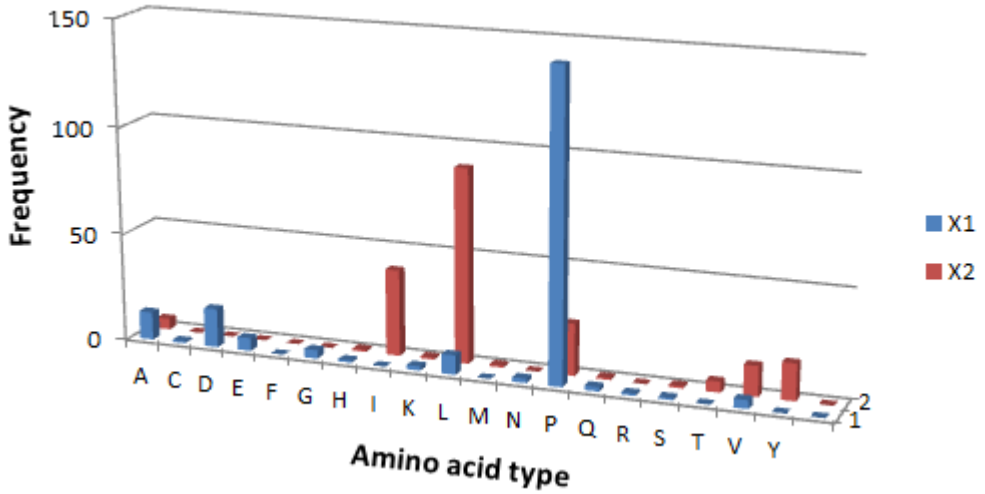


Figure 7-10. Amino acid distribution within the PxxP motif.

The analysis of the peptide sequences gives a general peptide consensus sequence of RPL**PPLP**PPP. This is an interesting observation as a large number of proline

residues are not typically found physiologically. This is likely because long polyproline peptides will typically form helical structures which would have difficulty associating with the SH3 domain.

7.3 Expression, purification and characterisation of a drug target region of PSD95

Construct design

A total of three constructs were produced, all of which were truncated forms of the full PSD95 sequence. The features of each construct are shown in Figure 7-11 overleaf.

7.4 Cloning, expression and purification of constructs A, B and C

7.4.1 Construct A of PSD95

The cDNA fragment corresponding to bases 417-724 of the Rattus norvegicus PSD-95 protein was amplified by polymerase chain reaction (PCR) from the full length sequence, using the following primers which were optimised for the Gateway® recombination cloning system (Invitrogen) ; forward primer (CCGAAAACCTGTATTTTCAGGGCTCAGGGACTGCATCCTTG) and reverse primer (GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATCAGAGTCTC TCTCGGGCTG). The resulting PCR product (Figure 7-12a) was purified and then further amplified using the previous reverse primer with the following forward primer (GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGAAGGAGATA

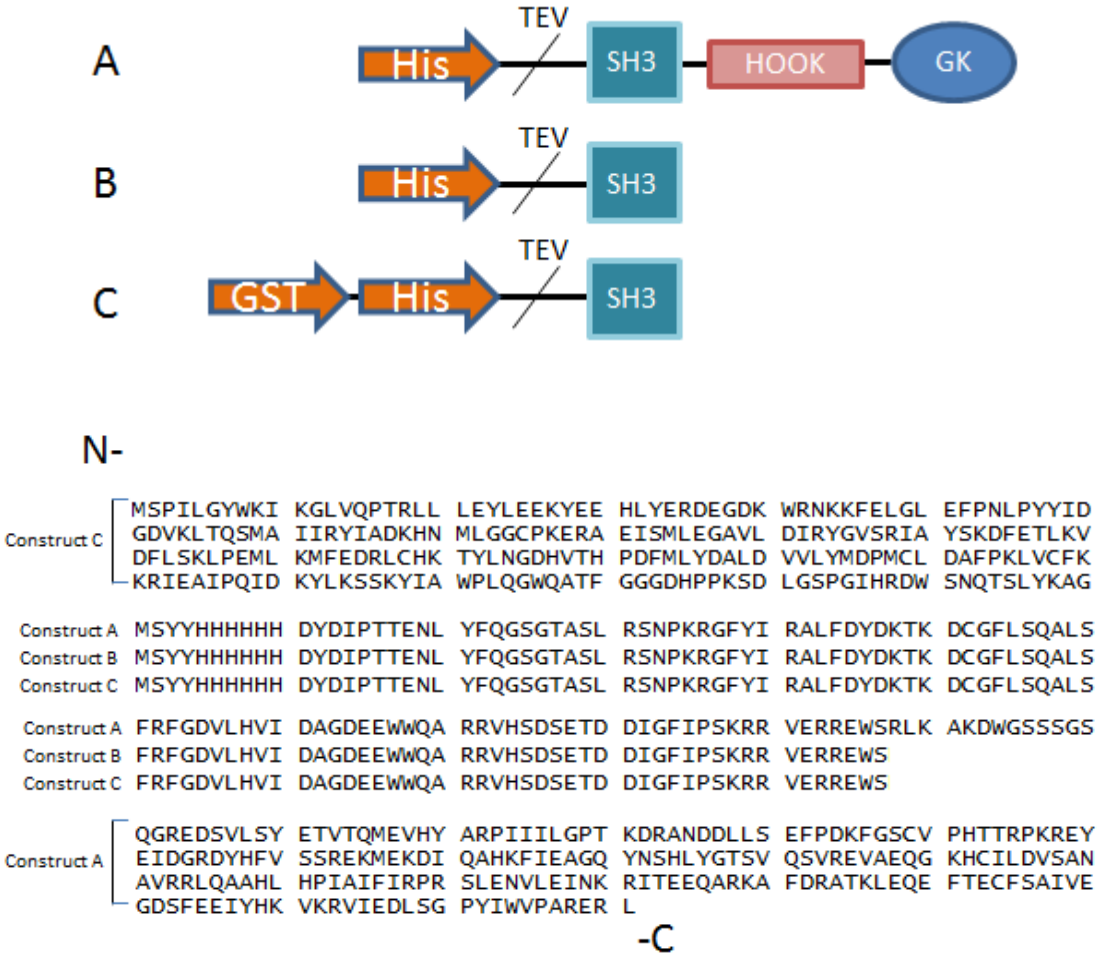


Figure 7-11. Construct design of druggable PSD95 recombinant protein. All constructs had the 3 PDZ domains at the N-terminus removed as these are unnecessary for testing of ligand binding. A TEV cleavage site was introduced to remove tags which were added for affinity purification. Alignment of all three sequences shows the larger N-terminal region prior to the SH3 domain in construct C contributed by GST. Construct A has a larger C-terminal region due to the presence of the Hook and GK domains.

TACATATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCA ATGACCGAAAACCTGTATTTTCAGGG). This PCR product (Figure 7-12b) was then cloned into the PDONR221 vector and subsequently shuttled into pDEST14. This correct sequence was verified by DNA sequencing (DNA Sequencing & Services, University of Dundee).

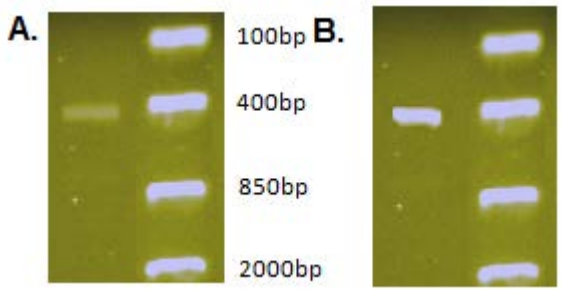


Figure 7-12 Cloning of SH3-Hook-GK construct. (a)PCR generation of fragment corresponding to residues 417-724 (lane 1). (b) PCR generation of TEV cleavable 6 His Tagged fragment ready for pDONR221 cloning.

Recombinant protein was expressed in *E. Coli* BL21(DE3) cells in LB media containing 50µg/ml carbenicillin. Cells were grown with shaking (250 RPM) at 37°C until the A600 reading was approximately 0.6 and expression induced by addition of 0.5mM IPTG. Growth was continued for 4 hours before cells were harvested by centrifugation at 12000 x g for 30 minutes. Cells were resuspended at 10% weight/volume in ice cold Lysis buffer (50mM sodium phosphate, 300mM NaCl, 10% glycerol) with excess protease inhibitor cocktail (Roche) and then lysed using a cell disruptor (Constant Systems) at 25KPSI. Cell lysate was centrifuged at 50,000xg for 1 hour and supernatant containing protein was passed through a 0.22 µm filter before being loaded onto a 5ml HiTrap IMAC FF column (GE Healthcare) which had been charged with Nickel and pre-equilibrated with Buffer A (Lysis buffer with 20mM Imidazole). A gradient of 0-100% Buffer B (Lysis buffer with 200mM Imidazole) was used to elute the protein.

Construct A protein was successfully expressed and purified to >95% purity from estimation of relative band intensity from coomassie stained SDS-PAGE gels (Figure 7-13). The final protein yield was approximately 20 mg per litre of culture.

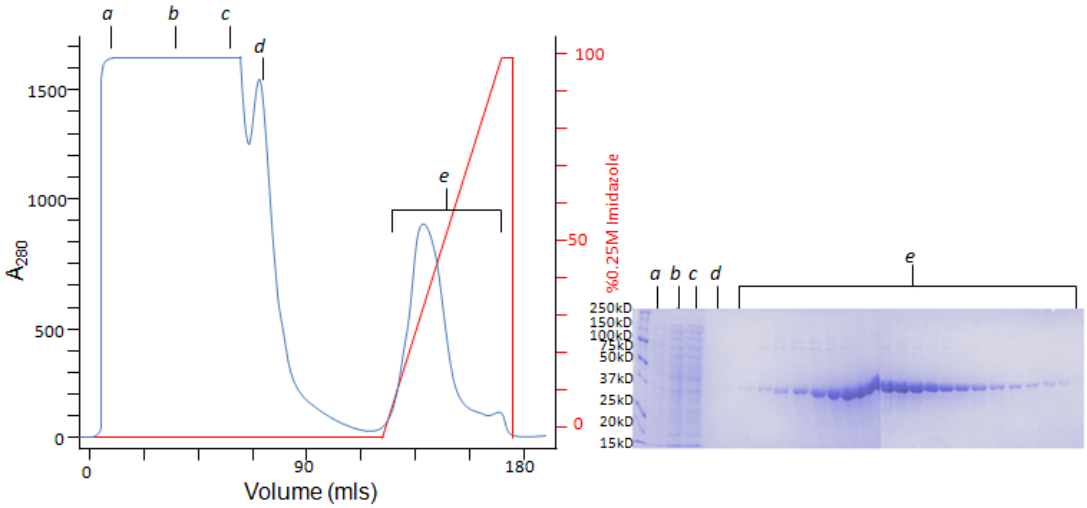


Figure 7-13. Purification of SH3-Hook-GK fragment of PSD95. The HiTrap IMAC FF column loaded with cell extract begins to elute PSD95 at approximately 10% 0.25mM Imidazole. Gel analysis of lettered fractions shows the recombinant protein purity is >95%.

7.4.2 Construct B of PSD95

The SH3-HOOK-GK DNA fragment was amplified using the same forward primer as described in 7.3.1 and a modified reverse primer (5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CTA TGA CCA CTC TCG TCG CTC GAC-3') which creates a DNA strand of 249 bases. This PCR product (Figure 7-14) was then cloned into the pDONR221 vector and subsequently shuttled into pDEST14. This correct sequence was verified by DNA sequencing (DNA Sequencing & Services, University of Dundee).

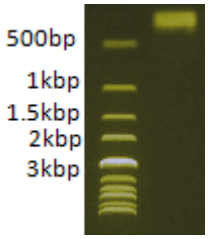


Figure 7-14. PCR product of SH3 domain alone was approximately 250bp in length.

Initial expression in BL21(DE3) showed the protein to be expressed in inclusion bodies. A number of different cell lines and expression times and temperatures were used to increase soluble protein however these were not successful. The C41(DE3) cell line at 37°C for 4 hours expressed the best insoluble yield (Figure 7-15), this insoluble protein was used in refolding experiments.

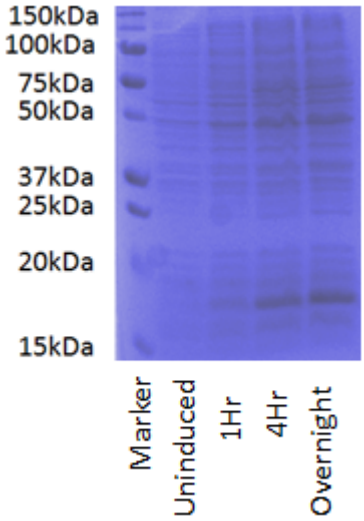


Figure 7-15. Expression trials with construct B showed highest expressions from C41(DE3) cells induced for 4 hours at 37°C.

Cell pellets were resuspended to 10% wt/vol in 25 mM NaH2PO4, pH 8.0 and passed once through a Constant Systems Cell Disruptor set to 22 kPSI, at 4˚C. Following this the lysate was centrifuged at 15,000 xg for 10 min at 4°C and the

pellet resuspended in 15 ml 25 mM NaH_2PO_4, 0.5 M NaCl, 0.5 % Triton X-100, pH 8.0 at 4˚C, and subjected to sonication (4 x 10 sec bursts). The sample was then centrifuged at 15,000 xg 10 min at 4 °C. The above was repeated twice and the resulting pellet resuspended in ~ 40 ml of 6 M guanidine hydrochloride, 25 mM NaH_2PO_4, 0.5 M NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0, homogenized in a dounce homogeniser, and left gently stirring for 2 hrs at 4˚C. The sample was then centrifuged at 15,000 xg for 15 min at 4 °C and passed sequentially through 0.45 μm and 0.22 μm filters. The filtered solution was then loaded onto a 5 ml nickel charged IMAC FF (GE Healthcare) column pre equilibrated in 6 M guanidine hydrochloride, 25 mM NaH_2PO_4, 0.5 M NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0, at 1 $ml.min^{-1}$. The column was washed for 10 column volumes of the same buffer followed by 10 col. vols. of 6 M guanidine hydrochloride, 25 mM NaH_2PO_4, 0.5 M NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0 at 1 $ml.min^{-1}$. A reverse gradient, over 50 column volumes, was then run with 25 mM NaH_2PO_4, 0.5 M NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol; 0.1% TX-100; 10% Glycerol, pH 8.0, followed by a further 5 column volumes of the same buffer at 1 $ml.min^{-1}$. The column was then washed with 10 column volumes of 25 mM NaH_2PO_4, 0.5 M NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol: 10% Glycerol, pH 8.0 at 1 $ml.min^{-1}$ and bound protein eluted with a gradient of imidazole from 20 mM to 500 mM over 20 column volumes at 1 $ml.min^{-1}$ as shown in Figure 7-16.

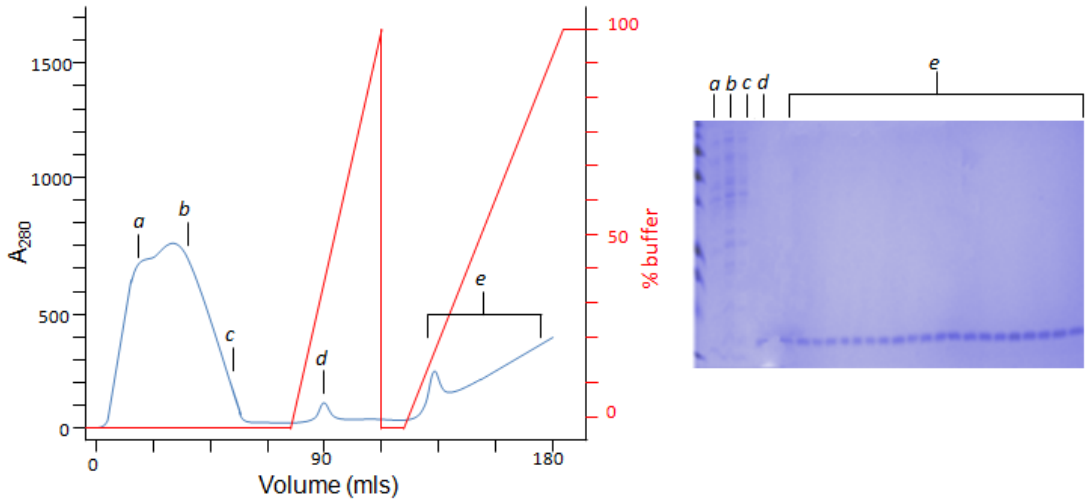


Figure 7-16. Refolding and purification profile of SH3 domain construct B.

The SH3 protein that was recovered from column refolding was >95% purity by coomassie gel analysis, however the protein precipitated shortly after elution. A number of methods were employed to keep the protein soluble including elution by pH instead of imidazole, refolding by dilution instead of on the IMAC column and use of additives. None of these were successful and further characterisation of the SH3 domain was not possible. To attempt to resolve this insolubility, construct C was created which is an SH3 domain with an N-terminal GST tag.

7.4.3 Construct C of PSD95

To create this construct the pDONR221 vector incorporating the SH3 domain sequence was used in the Gateway LR reaction with the pDEST15 vector. This correct sequence was verified by DNA sequencing (DNA Sequencing & Services, University of Dundee).

This construct was identical to construct B with the addition of an N-terminal GST tag. The tag was incorporated by Gateway recombinase reaction of the construct B

sequence into pDEST15 which has an N-terminal GST as part of the vector sequence.

Expression trials were conducted using three cell lines: Bl21(DE3), Rosetta Origami 2 and Origami 2 (DE3). 10ml cultures were setup for each cell line and cell samples were taken before induction, 4 hours after induction and then after overnight induction. Cultures were grown at 37 °C with shaking at 250RPM and 50µg/ml carbenecillin.

Expression trials showed that Rosetta Origami 2 and Origami 2 (DE3) cells had much slower growth rate than the BL21 (DE3) cells. However after overnight induction production of SH3-GST was considerably higher in the Origami 2 (DE3) cells. Protein production in the most favourable conditions is shown if Figure 7-17.

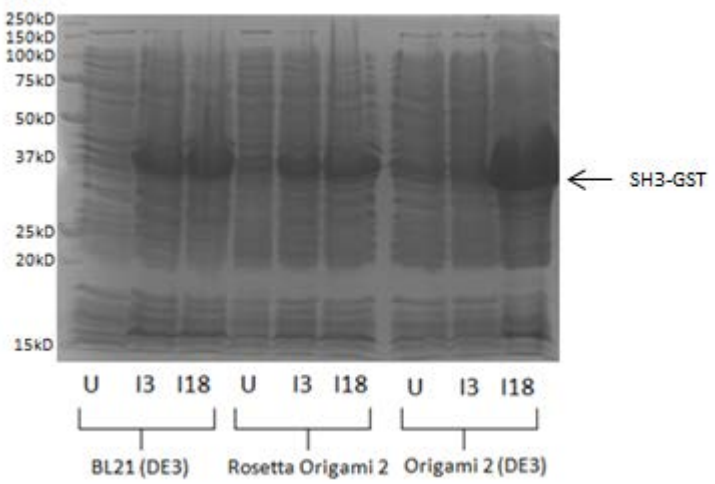


Figure 7-17. Whole cell pellet SDS PAGE gel of construct C overexpression trials. Gel shows uninduced (U), Induced for 3 hours (I3) and induced for 18 hours (I18) results.

Culture volumes were scaled up to 1L and cells were grown until an O.D of 0.6 was reached then induced with 0.5mM IPTG for 4 hours. Cells were harvested by

centrifugation at 12000 x g for 30 minutes. Cells were resuspended at 10% weight/volume in ice cold Lysis buffer (50mM sodium phosphate, 300mM NaCl, 10% glycerol) with excess protease inhibitor cocktail (Roche) and then lysed using a cell disruptor (Constant Systems) at 25KPSI. Cell lysate was centrifuged at 50,000xg for 1 hour and supernatant containing protein was passed through a 0.22 µm filter before being loaded onto a GSTrap column and eluted using a gradient of 0-10mM reduced glutathione 50mM Tris HCl, pH 8.

The addition of the GST tag did increase the solubility of the protein, however there was a large number of other contaminating proteins which co-purified with construct C (see Figure 7-18).

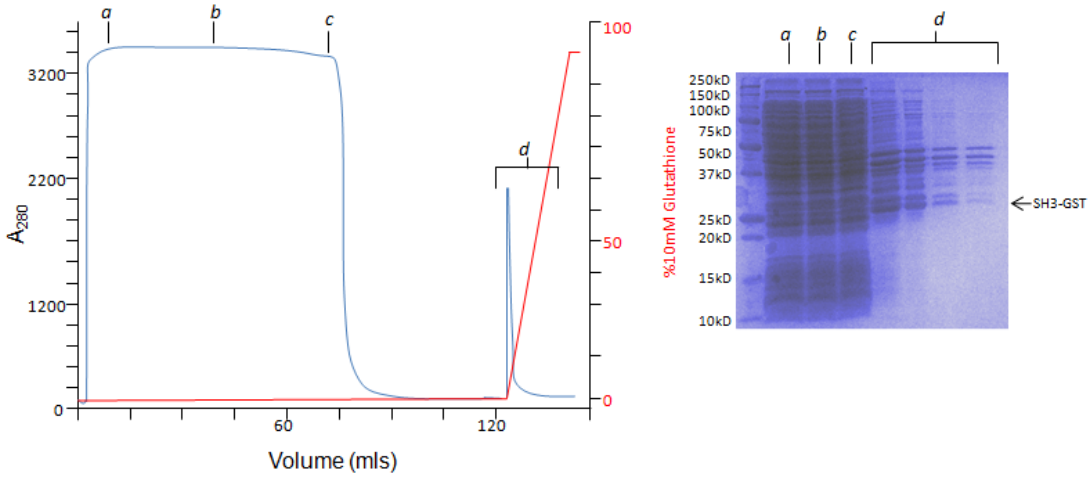


Figure 7-18. Purification trace of SH3-GST and SDS-PAGE analysis. Elution of the protein was rapid following low concentration of reduced glutathione. This low binding to the GSTrap column caused high contamination of the sample.

A number of methods were used to try to reduce the contamination; the elution gradient was made shallower, eluted fractions were desalted into PBS and then re-run over the GSTrap column and ion exchange was also attempted. None of the methods

tested were able to significantly decrease the contamination and further characterisation with the construct was not possible.

7.4.4 Discussion of PSD95 constructs and purification methods

Construct A was designed to remove the three N-terminal PDZ domains as they were not required for binding assay. Construct B was designed to probe the stability of the SH3 domain without the strands which are contributed by the Hook and GK domains. This deletion significantly compromised the stability of the domain and caused the protein to be expressed within inclusion bodies, attempts to refold the protein into a soluble form were unsuccessful. Construct C was designed to increase the solubility of the truncated SH3 domain by addition of the large GST tag[32] which was successful; however difficulty was encountered in achieving separation from contaminating proteins. Therefore construct A was the only protein used in further characterisation studies.

7.5 Characterisation of PSD95 construct A

7.5.1 Gel filtration analysis of construct A

Gel filtration studies were carried out using a Superdex 75 10/300 GL column at a flow rate of 0.5ml/min. The column was equilibrated with 50mM sodium phosphate (pH 8), 300mM NaCl, 10% Glycerol at 4°C. Molecular weight calibration was performed using the following standards: Aprotinin (6500Da), Ribonuclease A (13700Da), Carbonic Anhydrase (29000Da), Ovalbumin (43000Da) and Conalbumin (Da75000).

Concentrated SH3-HOOK-GK protein was applied to a Superdex 75 10/300 GL column and eluted with sodium phosphate buffer (Figure 7-19). The results showed that the majority of the protein was a monomer in solution with a small percentage existing as a dimer. This is in agreement with the idea that the SH3-HOOK-GK modules found in MAGUK proteins can dimerise to form higher order structures[33].

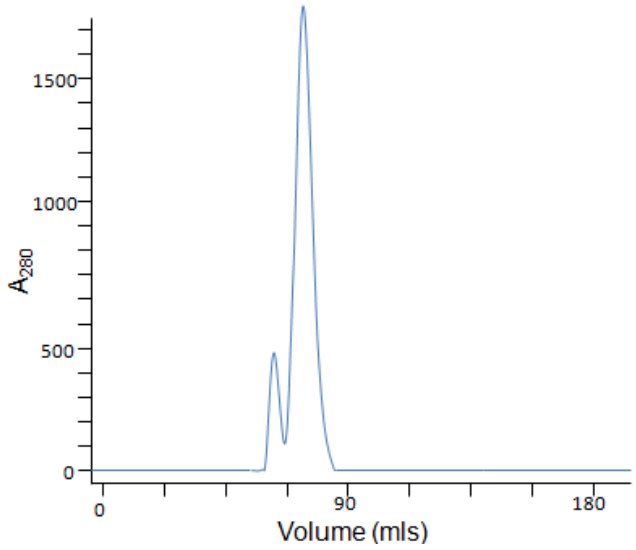


Figure 7-19. Gel filtration analysis of construct A protein reveals the protein is primarily in a monomeric state.

7.5.2 Thermal denaturation of construct A

In a thermal denaturation experiment the target protein is heated over a temperature range which completely unfolds the protein. A compound which binds to the protein and emits a fluorescent signal as the unfolding continues is included in the assay and allows measurement of the change in fluorescence (and therefore melting temperature). This provides a rapid method to assess how well folded the protein is. Furthermore thermal denaturation assay is recognised to be a useful technique in high throughput testing of potential drug compounds[34]. In addition there is experimental evidence to suggest that the data obtained from such studies is

comparable in quality to that obtained from Isothermal Calorimetry (ITC) even at low (<50) µM K_d[35] . The assay is performed in the same way as a standard thermal denaturation experiment but with ligand molecules added, thereafter shifts in unfolding temperatures are monitored. The advantages of this technique over more traditional methods is that it is easily adaptable to high throughput screening since the assays can be set-up in 96 well plates and no information is required on the target protein's binding site.

When screening compounds the expected result for a positive hit is an increase in melting temperature as a result of compound binding contributing stabilising effects to the protein. However this is not always the case, there are occasions where a decrease in melting temperature in response to ligand binding can occur. This has been attributed to the ligand binding to an unfolded state of the protein[36].

Protein concentration screening

A protein concentration range of 0.5-8 µM with 10X Sypro orange (Invitrogen) was used to assess the optimum concentration to obtain a low signal to noise ratio. Sample volume was 25 µl protein and 25 µl 20X Sypro orange in buffer (50 mM $NaPO_4$ pH 8, 300 mM NaCl and 10% glycerol). All experiments were carried out in a thin walled 96 well plate (Bio-rad) closed with sealing tape and measurements were made using a Cycler iQ Real Time PCR Detection System (Bio-Rad). Each sample was measured in triplicate and the averages were calculated.

The wavelengths for excitation and emission were 490 nm and 575 nm respectively. Fluorescent readings were taken between 20 °C and 80 °C in 1 °C increments after each temperature had been held for 90 seconds.

The protein melt concentration results showed that the protein had unfolded completely by 61 °C (Figures 7-20). From these tests it was decided that 4 µM protein was an optimal concentration given the signal strength and low error.

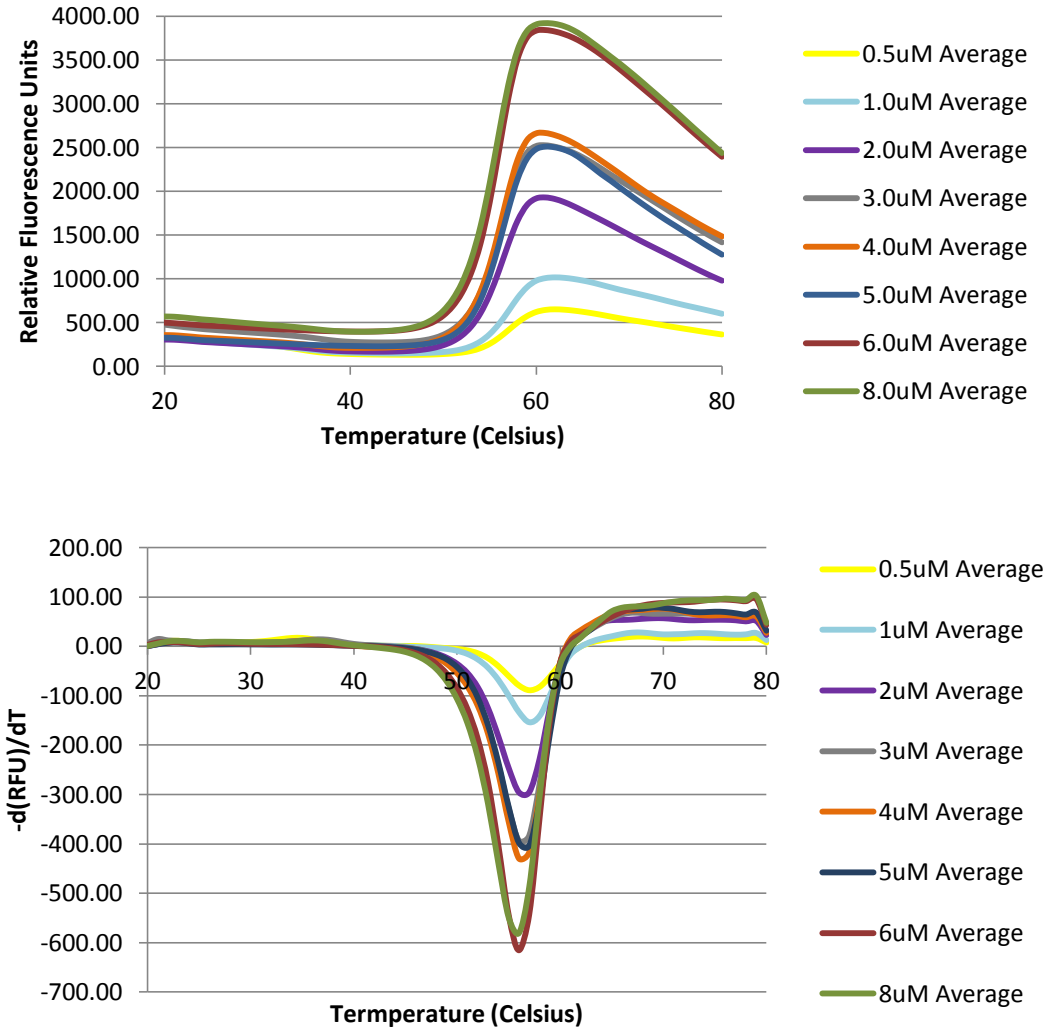


Figure 7-20. Protein concentration effects on thermal melting.

DMSO concentration screening

In the majority of cases small molecule stock solutions are made using DMSO as the solvent. To understand what effects DMSO concentration can have on the protein stability a broad range of DMSO concentrations 0.5-21% were tested to establish the

maximum concentration of DMSO in our experiments to limit the effect that DMSO had on our results.

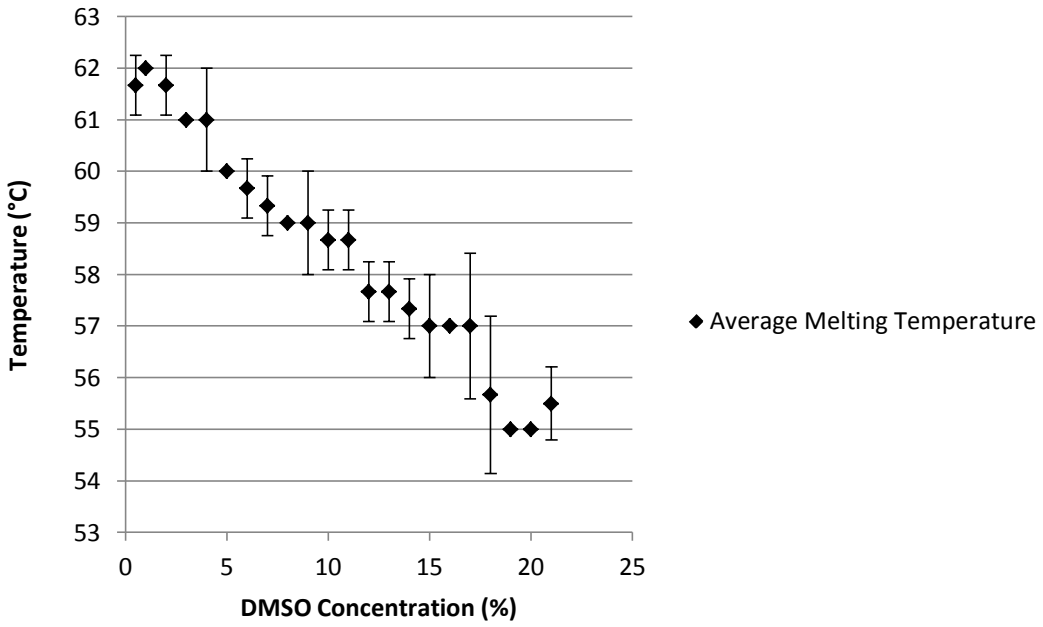


Figure 7-21. Effects of increasing DMSO concentration on protein stability in thermal denaturation studies. In low DMSO concentration there is an observed increase in the protein melting temperature

From Figure 7-21 it can be seen that as DMSO concentration increases the melting point of the protein decreases. The exception to this is in the 0.5-2% region where there is an observed increase in melting temperature with a maximal increase over the mean melting temperature by 1 °C. It may be the case that DMSO is able to bind to parts of the protein structure non-specifically which causes slight increases in protein stability. Eventually a tipping point is reached where non-specific DMSO binding induced stability is offset by the increasing disruption of secondary structure which is caused by higher DMSO concentration[37]. A decrease in melting temperature by 6 °C was caused at 19 %, beyond this point no further change was

found up to a concentration of 21 %. From these results it was decided that a final DMSO concentration of no more than 5% would be used in compound assay. This is a useful concentration as it doesn't cause a large increase or decrease in protein melting temperature whilst still allowing compounds to retain solubility in solution.

7.5.3 Circular dichroism analysis of construct A

All spectra were recorded on a nitrogen flushed Jasco J-810 spectropolarimeter. Protein was dialysed into 50 mM $NaPO_4$ (pH 8) and 50 mM KF at a concentration of 2.4 µM and used with a cell of 0.02 cm pathlength. Data was recorded from 250 to 190 nm with a 1m bandwidth and 10 nm min^{-1} scan speed. Secondary structure was deconvoluted using the SELCON3 method on the DICHROWEB server[38].

To confirm that the protein was well folded 2.4 µM protein was used in circular dichroism analysis (Figure 7-22). The results showed that the protein was folded correctly and the approximate alpha helix content (39%) was close to that of the reported crystal structure (31%).

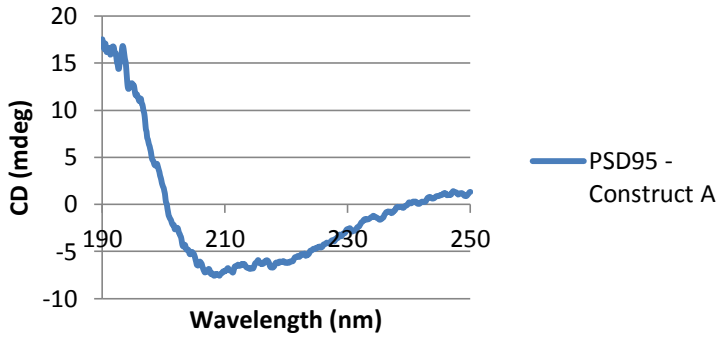


Figure 7-22. Analysis by circular dichroism shows that the recombinant protein is well folded.

7.5.4 Discussion on the characterisation of construct A

From the characterisation studies it can be concluded that the protein is well folded with structure which is in close agreement with the literature. This study highlights that though the protein can exist as a dimer this does not affect the thermal stability of the protein, a single melt peak is consistently observed irrespective of increasing protein concentration in the range 0.5-8 µM.

The melting profile can be influenced by the presence of DMSO solvent, literature studies show that DMSO causes instability in proteins[39]. Whilst this was also observed in this study, at concentrations below 5% there was an increase in stability. This was a useful observation as this construct was then utilised for high throughput small molecule virtual screening, given the heavy use of DMSO as an organic solvent for small molecule libraries.

7.6 LIDAEUS docking into the PSD95 SH3 domain

7.6.1 Defining the site points of PSD95 SH3-HOOK-GK

The area targeted for LIDAEUS docking was defined in a similar way to GnRH-R as described in Chapter 2. Briefly, a peptide molecule was used to define the area on the protein surface that was used to generate an energy map which could then be submitted to LIDAEUS. The site points that were used in the docking study are shown in Figure 7-23.

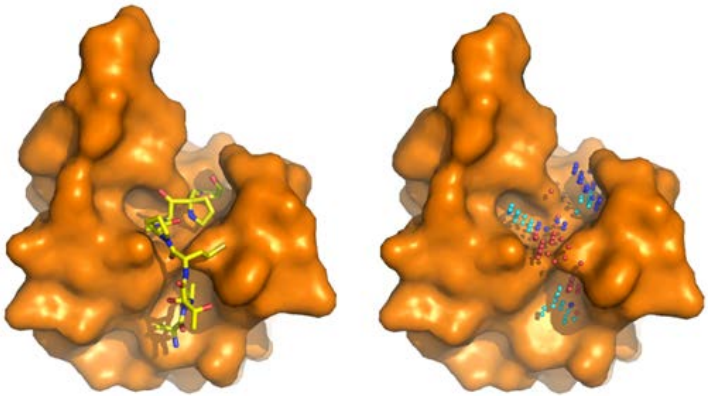


Figure 7-23. Determination of site points for LIDAEUS docking experiments. (Left) A peptide molecule is orientated along the expected binding region of the SH3 domain to define the spacial area used to generate an energy map. (Right) The site points which were used in LIDAEUS docking. Cyan spheres represent hydrophobic sites, blue represents hydrogen bond donors and red represents hydrogen bond acceptors.

7.6.2 LIDAEUS results using multiple conformers and flexible docking

In this docking experiment the workflow was changed to accommodate an AutoDock step after a multi-conformer version of the EDULISS database was docked into the protein by LIDAEUS. Previously the EDULISS database in use was a single conformer version. The top 5000 compounds from the LIDAEUS docking were re-docked and re-ranked by AutoDock (using the same method as for GnRH-R in Chapter 2) of which the top 100 were used in clustering (Figure 7-24).

From the clustered compounds, 6 were chosen for assay. These were supplied by ChemBridge and were easily available. However they still represented a high scoring section with 4 of the compounds chosen residing in the top 20. The structure, LIDAEUS ID, Chembridge ID and Autodock scores of the selected compounds are shown in Table 7-2.

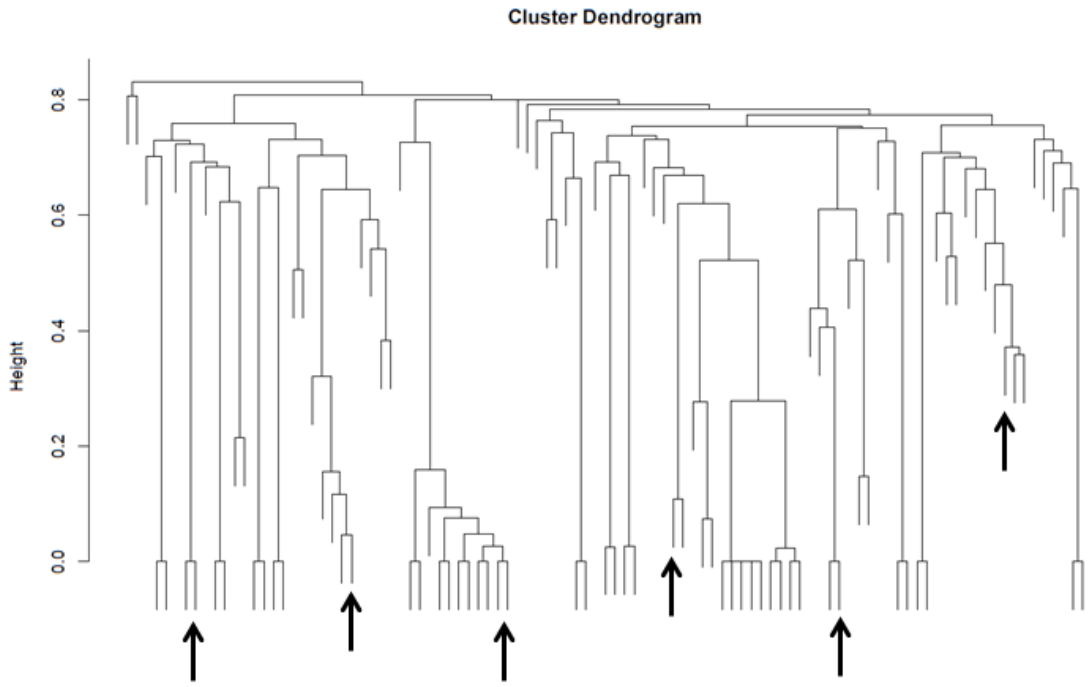
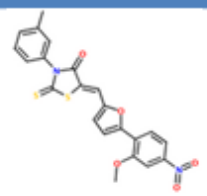
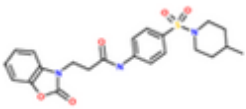
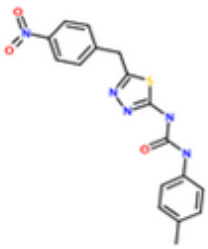
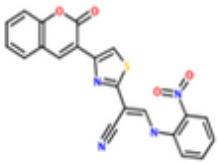
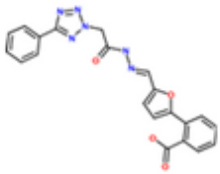
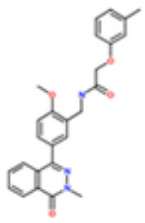


Figure 7-24. Compound clustering results for PSD-95 from LIDAEUS. Position of compounds selected for assay are indicated by black arrows.

The suggested potential interaction sites between each compound and the PSD95 SH3 sites are shown in Figure 7-25. Examination of the docked structures shows three compounds which avoid binding through the conserved SH3 ASP residue and three structures which dock at least part of the molecule into the ASP conserved region. Ligand association to more than one part of the PxxP binding region is useful because it allows greater flexibility when selecting for binding, specifically for PSD95 as well as opening up a broader selection of scaffolds which can be modified for better affinity if they are found to be active compounds.

Table 7-2. Selected Chembridge compounds to be assayed against PSD95.

Structure	Predicted affinity (kcal/ mol)	Predicted affinity (nM)	EDULISS ID	Chembridge ID
	-8.82kcal/mol	342.1	SPH1-243-918	6786900
	-8.71kcal/mol	411.6	SPH1-396-579	7940094
	-8.34kcal/mol	771.4	SPH1-286-069	7202320
	-8.28kcal/mol	852.0	SPH1-077-118	5620127
	-8.18kcal/mol	1010	SPH1-087-855	5676364
	-8.00kcal/mol	1360	SPH1-328-947	7643136

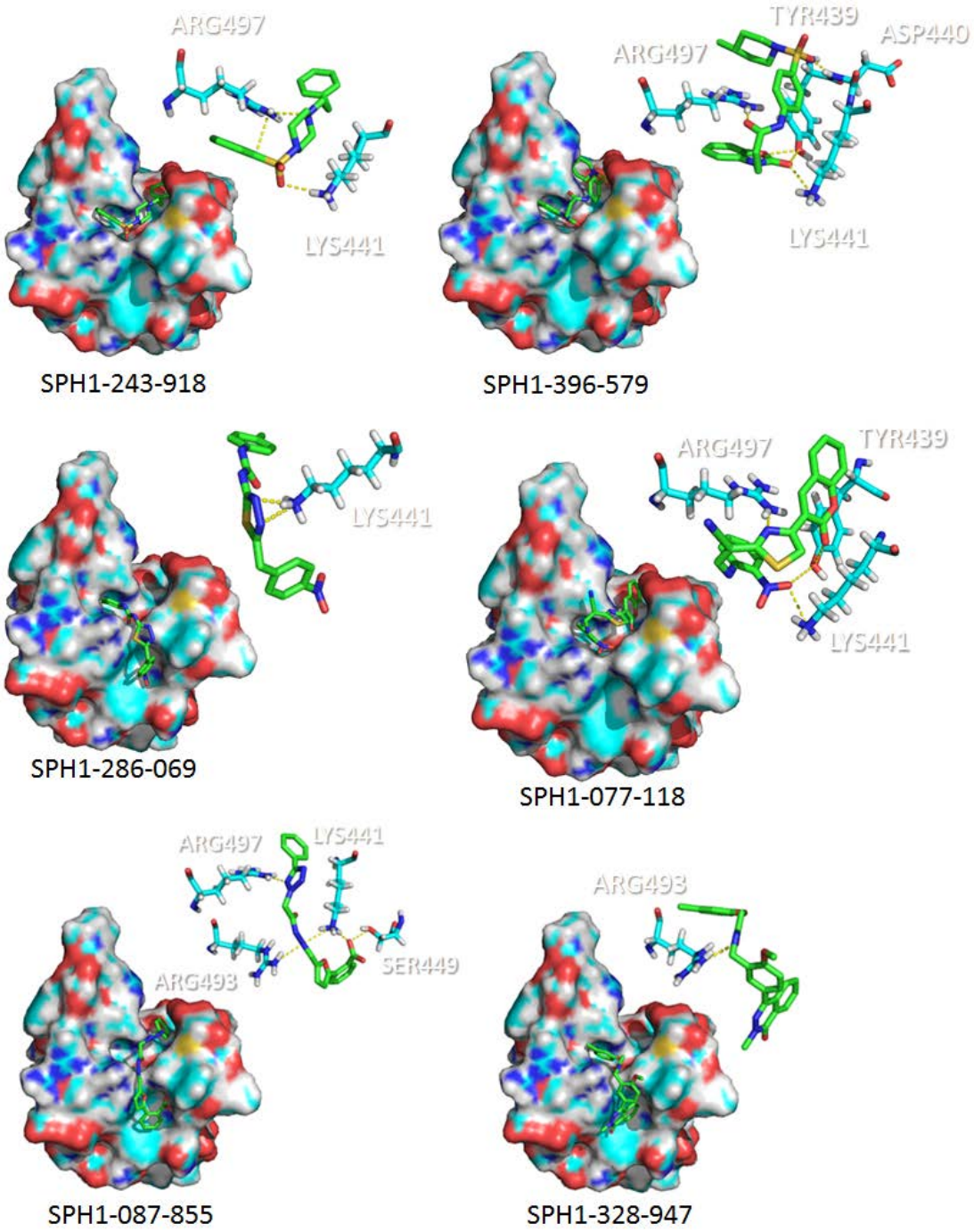


Figure 7-25. Potential binding interactions of 6 LIDAEUS conformers docked into the SH3 domain of PSD95.

7.7 Assay development for the in vitro testing of small molecule compounds from virtual screening

A key component of the drug development workflow is the ability to test the compounds that are selected from virtual screening. In this part of the study the use of two assay techniques are assessed; the first is a fluorescence based assay which utilises intrinsic tryptophan fluorescence, the second is a competition binding assay which uses a fragment of a protein which has been shown to bind to the SH3 domain.

7.7.1 Using intrinsic tryptophan fluorescence as an assay technique for LIDAEUS compounds

Construct A contains four tryptophan residues, the W470 residue which is well conserved among SH3 domains is a requirement of binding in the specificity pocket[40]. As this tryptophan is located directly in the targeted binding site for the small molecule ligands (See figure 7-26), then changes in fluorescence intensity can indicate interactions between small molecules in environments close to the tryptophan residues.

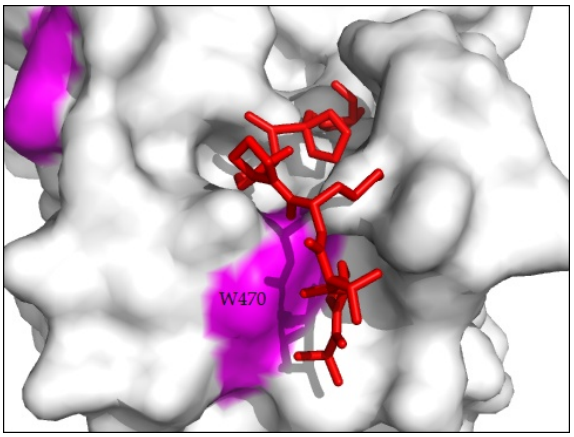


Figure 7-26. The SH3 domain of PSD95 contains a tryptophan residue (magenta) in the specificity pocket which is exposed to the surface. An example peptide typical of an SH3 binding molecule is shown in red.

Fluorescence measurements were taken using a Fluoromax-3 spectrofluorometer (Horiba Jobin Yvon) with a 3.5nm slit width. Total sample volume was 600 µl in a 1 cm path length quartz cuvette at 4 °C. The excitation wavelength was 295 nm and the emission range was from 320-450 nm in 0.5 nm increments. Readings were taken in triplicate and then averaged. All readings were corrected for using buffer blanks.

All compounds were dissolved in 100% DMSO, however compound #6786900 was insoluble in 100% DMSO and was not used in further assays. The compounds themselves were tested first to establish their native fluorescence by excitation at 295nm and scanning for emission from 320 to 450 nm with a slit width of 5nm (Figure 7-27).

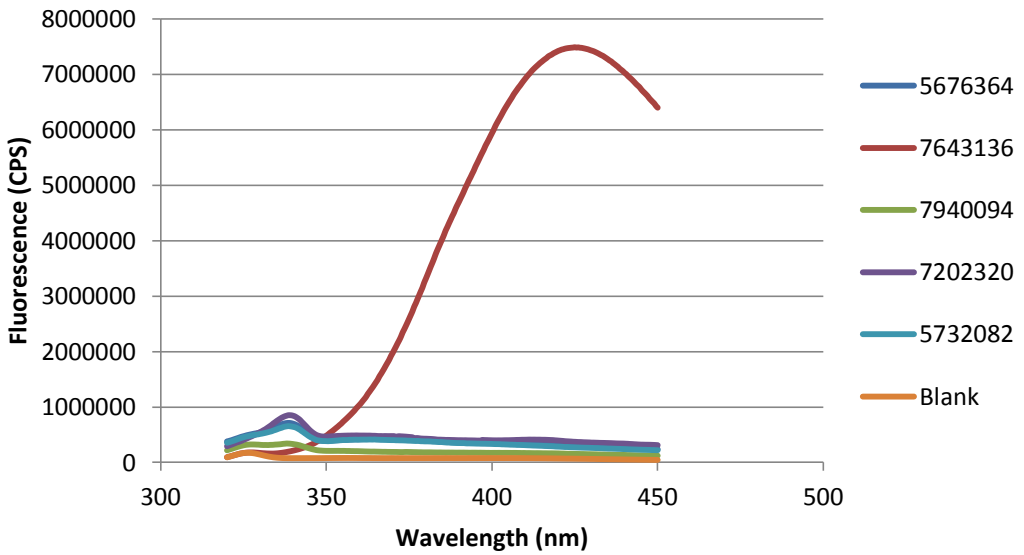


Figure 7-27. Intrinsic fluorescence from LIDAEUS compounds selected for assay. 20% DMSO was used as the blank.

All compounds exhibited some fluorescence, however compound # 7643136 was so fluorescent that it gave a reading higher than that which the spectrofluorometer can

accurately read; therefore this compound was removed from further assay study. The remaining four compounds were tested with 0.5 µM of construct A protein at a concentration of 125 µM (Figure 7-28). The changes in fluorescence as a result of compound incubation were quite dramatic. Compound #5676364 caused a large red shift of the peak by approximately 60 nm. None of the other compounds caused such a large change, however they did cause the peak to change shape and sharpen with a blue shift by approximately 10 nm indicating an increased non-polar environment.

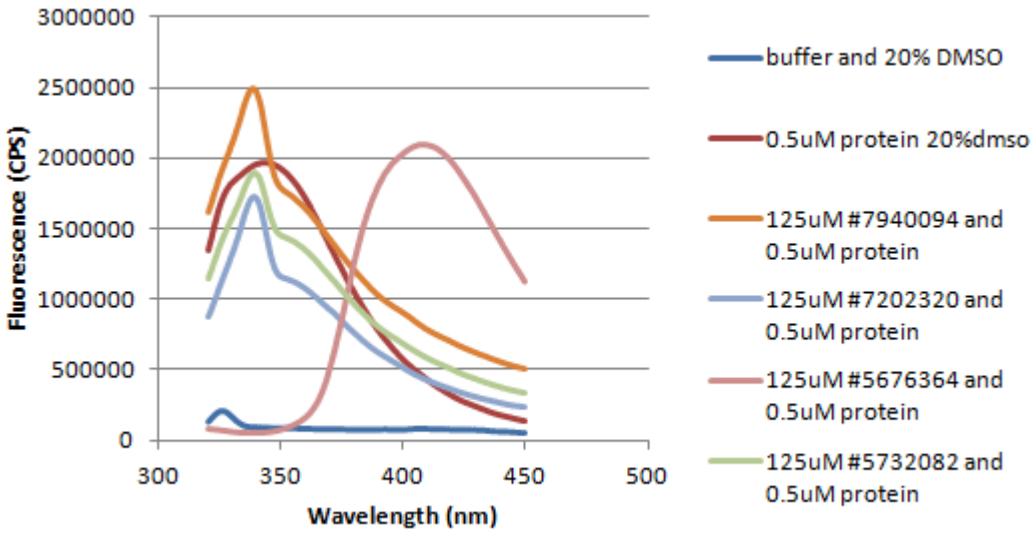


Figure 7-28. Effects of compound incubation on tryptophan fluorescence.

Large red shifts in tryptophan fluorescence spectra are typically observed in experiments involving protein interactions with chaotropic agents, high concentrations of urea or GnHCl unfold proteins causing increased solvent exposure to the tryptophan residues[41]. However these wavelength changes are usually less than 20 nm. The change in protein structure which has resulted in this large red shift caused by compound #5676364 is an area that would be interesting for further study.

The remaining compounds caused a similar change in peak shape however they differed in respect to how they affected the fluorescence intensity. Compounds

#5732082 and #7202320 both acted as fluorescence quenchers, the observed quenching may be attributed to a dynamic or static effect. Which effect is causing the reduction in signal intensity is unclear; as the tryptophan is surface exposed the formation of a complex (static) or just small molecule/tryptophan collision (dynamic) are possible.

The increase in fluorescence that was observed from #7940094 is in contrast with the expected result, namely a quenching effect as a result of direct ligand/small molecule association as was observed for compounds #5732082 and #7202320. However the increase in fluorescence suggests that there is an increase in hydrophobic environment directly surrounding a tryptophan residue[42], possibly caused by conformational changes induced by small molecule association elsewhere on the molecule.

7.7.2 Using competition binding for LIDAEUS compound testing

The peptide sequence of the PI(3)K-C2α fragment which was shown to bind to the SH3 domain of PSD95 had not been published prior to the beginning of assay development with the truncated PSD95 protein. At this point in assay development there were four known proteins which could associate with PSD95 SH3 domain. These were KA2[43], Pyk2[44], Huntingtin[45] and the GK domain of PSD95[46]. The aim was to obtain a peptide sequence which could bind to the SH3 domain and use it in a competitive assay. Two peptide sequences were tested; MLKRLVPPPPKRMIYand FITC – LLDDPGPPGGP RPQAPTP. The MLKRLVPPPPKRMIY sequence was obtained as a result of private communication with S Grant at the Sanger Institute. This was tested by thermal denaturation assay,

however no significant change in melting point was observed even at a 50:1 peptide : protein ratio.

The FITC – LLDDPGPPGGPRPQAPTP sequence was obtained from part of the kainite 2 subunit which was found to bind to the SH3 domain via a proline rich sequence near the C-terminus[47]. The FITC chromophore was added to try to establish binding activity via fluorescence polarisation. This also yielded a negative result.

Rather than continue with peptide sequences the C-terminal fragment of the Kainate 2 receptor was ordered from Genscript and subsequently cloned into the pDEST14 vector for expression and purification.

7.7.2.1 Cloning KA2 C-terminal construct

The cDNA sequence corresponding to residues 861-979 of the Glutamate receptor, ionotropic kainate 5 from Rattus norvegicus was purchased (Genscript) and cloned into the pDEST14 gateway expression vector using the forward primer (CCGAAAACCTGTATTTTCAGGGCATGGAAAACATTGGCG GTATCTT) and reverse primer (GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTATTCAT GTTCGGTC AGTTC).

The cDNA sequence corresponding to residues 861-979 of the KA2 receptor was amplified by PCR with primers incorporating the Gateway recombinase sites (Figure 7-29a) and then finally re-amplified using primers containing the gateway recombinase sites with a TEV cleavable histidine tag (Figure 7-29b).

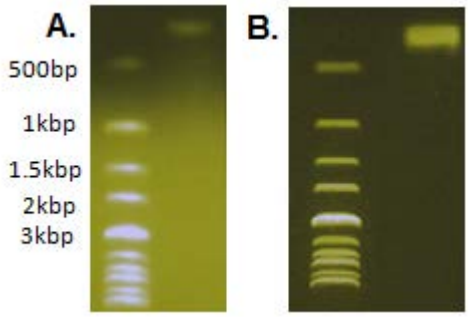


Figure 7-29 Cloning of SH3/GST domain construct. (a)PCR generation of fragment corresponding to residues 861-979 (lane 1). (b) PCR generation of TEV cleavable 6 His Tagged fragment ready for pDONR221 cloning.

7.7.2.3 Overexpression and purification of the KA2 C-terminal construct

To optimise the protein yield expression trials were performed using two cell lines with three different growth media and two concentrations of IPTG as shown in Figure 7-30.

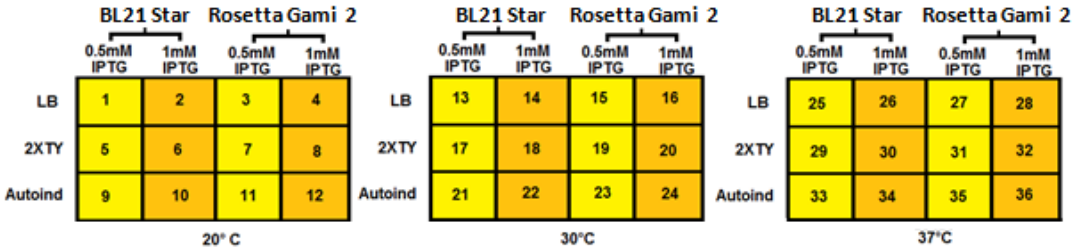


Figure 7-30. Optimising expression of KA2 N-terminal fragment. Recombinant protein expression was tested in 36 conditions by varying cell line, growth media, IPTG concentration and temperature.

Cell pellets were resuspended to 10% wt/vol in 25 mM NaH2PO4, pH 8.0 and passed once through a Constant Systems Cell Disruptor set to 22 kPSI, at 4˚C. The cell lysate was centrifuged at 15,000 xg for 10 min at 4°C. The pellet was resuspended in 8M urea and applied to a nickel charged IMAC column at a flow rate of 1ml/min. A gradient was applied of 6M-1M urea in 50mM Sodium phosphate

(pH 7.5) 10mM Tris-(pH 7.5), 20% glycerol followed by a 5 column volume wash with 50mM Imidazole in PBS. A gradient from 20mM to 500mM imidazole in PBS then followed to elute the target protein.

Expression was optimised using the BL21Star and Rosetta Gami 2 cell lines in addition to varying IPTG concentration, temperature and media type. As the overexpressed protein was contained within insoluble inclusion bodies the insoluble fraction was resuspended in urea prior to loading onto the column. After the refolding step a standard imidazole gradient in PBS was suitable for the elution of the protein (Figure 7-31). The refolded KA2 fragment was stable in the PBS buffer and no precipitation was observed.

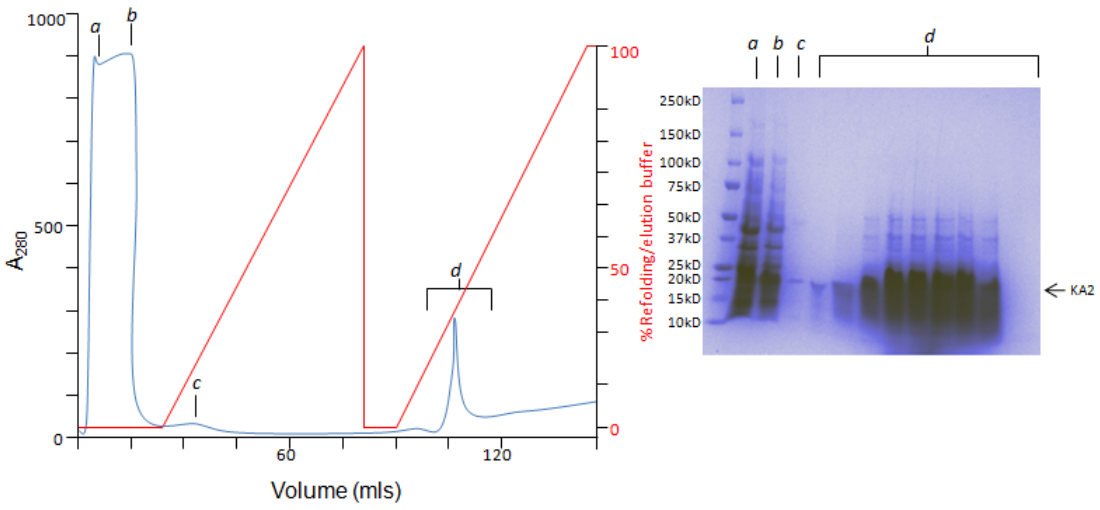


Figure 7-31. Purification of KA2 N-terminal fragment. Purification required an initial urea gradient to refold the protein followed by an imidazole gradient to elute. Very little protein was lost during the refolding and wash steps.

7.7.2.4 Assessing interaction of the KA2 fragment and construct A by native gel

To test if the fragment was able to bind to construct A containing the SH3-HOOK-GK domains a native gel was run (Figure 7-32)

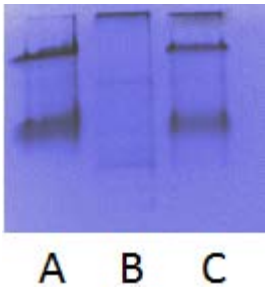


Figure 7-32. Assessment of Construct A binding to KA2 subunit using a native gel. (a) Construct A. (b) refolded KA2 subunit. (C) Construct A and refolded KA2 subunit.

The results of the native gel suggested that there is very little interaction between the KA2 fragment and construct A. Possible reasons for this include the KA2 fragment being folded incorrectly during the on-column refolding step. Due to time constraints further investigation and development of this assay method were not possible.

7.8 Conclusions and future work

The crystal structure of the SH3-HOOK-GK domains of PSD95 was available for the virtual screening work for this protein which contained a non-canonical SH3 domain. While the SH3 domain is non-canonical, literature evidence shows that the domain is still able to facilitate traditional proline rich sequence association. Structural study using STP indicated that this was still likely to be via the selectivity pocket and surrounding area despite the presence of an alpha helix in the binding region. Following protein purification, thermal denaturation studies showed that the protein melted at 61°C. This melting temperature was sensitive to increasing DMSO

concentration which after initially causing an increase in melting point caused a gradual decrease.

Small molecule ligands were identified using a database mining approach and the development of an assay was attempted. The use of tryptophan fluorescence was particularly suited to this study given the location of a highly conserved tryptophan in the specificity pocket which was a target for ligand docking. All molecules tested affected the tryptophan fluorescence profile; the most interesting change was from compound 5676364 which caused a large red shift of the peak. This suggested that the compound may have chaotropic properties, however a literature search of molecules of similar structure did not confirm this.

In future work it would be useful to incubate the PSD95 construct A protein with compound 5676364 and then measure the circular dichroism spectra to look for unfolding activity caused by the small molecule.

The activity of the other compounds in the study was varied in that there were increases and decreases observed which were compound dependent. It is not possible to confirm if these changes are caused as a result of binding to the PxxP binding site however competition assay with a peptide sequence FPLWKLPGFPNRMVLG from the PI(3)K-C2α signalling molecule could give more insight. Furthermore if the peptide sequence was tagged with a fluorophore like fluorescein or FITC then fluorescence polarisation could be used as a possible rapid assay method, this could be scaled for use in 96 well or 384 well plate formats with an appropriate plate reader.

7.9 References

1. Garry, E.M., et al., *Neuropathic Sensitization of Behavioral Reflexes and Spinal NMDA Receptor/CaM Kinase II Interactions Are Disrupted in PSD-95 Mutant Mice.* Current Biology, 2003. **13**(4): p. 321-328.

2. Arbuckle, M.I., et al., *The SH3 domain of postsynaptic density 95 mediates inflammatory pain through phosphatidylinositol-3-kinase recruitment.* EMBO Rep. **11**(6): p. 473-478.

3. Legrand, E., et al., *Sciatica from disk herniation: Medical treatment or surgery?* Joint Bone Spine, 2007. **74**(6): p. 530-535.

4. Alexandropoulos, K., G. Cheng, and D. Baltimore, *Proline-rich sequences that bind to Src homology 3 domains with individual specificities.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(8): p. 3110-3114.

5. Harkiolaki, M., et al., *Structural basis for SH3 domain-mediated high-affinity binding between Mona/Gads and SLP-76.* EMBO J, 2003. **22**(11): p. 2571-2582.

6. Ball, L.J., et al., *Recognition of Proline-Rich Motifs by Protein–Protein-Interaction Domains.* Angewandte Chemie International Edition, 2005. **44**(19): p. 2852-2869.

7. Bar-Sagi, D., *Transmembrane Signaling Protocols (Methods in Molecular Biology)*. 2006: Humana Press. 18-19.

8. Guerois, R. and M.L.d.l. Paz, *Protein Design: Methods and Applications (Methods in Molecular Biology)*. 2006: Humana Press. 216-217.

9. Ferreon, J.C. and V.J. Hilser, *Ligand-induced changes in dynamics in the RT loop of the C-terminal SH3 domain of Sem-5 indicate cooperative conformational coupling.* Protein Science, 2003. **12**(5): p. 982-996.

10. Migliaccio, A., et al., *Inhibition of the SH3 domain-mediated binding of Src to the androgen receptor and its effect on tumor growth.* Oncogene, 2007. **26**(46): p. 6619-6629.

11. Oneyama, C., H. Nakano, and S.V. Sharma, *UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug.* Oncogene, 2002. **21**(13): p. 2037-2050.

12. Sharma, S.V., et al., *UCS15A, a non-kinase inhibitor of Src signal transduction.* Oncogene, 2001. **20**(17): p. 2068-2-79.

13. Weng, Z., et al., *Structure-function analysis of SH3 domains: SH3 binding specificity altered by single amino acid substitutions.* Molecular and Cellular Biology, 1995. **15**(10): p. 5627-5634.

14. Politou, A.S., et al., *SH3 in muscles: solution structure of the SH3 domain from nebulin.* Journal of Molecular Biology, 1998. **276**(1): p. 189-202.

15. Jozic, D., et al., *Cbl promotes clustering of endocytic adaptor proteins.* Nat Struct Mol Biol, 2005. **12**(11): p. 972-979.

16. Schmidt, H., et al., *Solution Structure of a Hck SH3 Domain Ligand Complex Reveals Novel Interaction Modes.* Journal of Molecular Biology, 2007. **365**(5): p. 1517-1532.

17. Lee, C.H., et al., *A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein.* EMBO, 1995. **14**(20): p. 5006–5015.

18. Rickles, R.J., et al., *Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries.* EMBO, 1994. **13**(23): p. 5598–5604.

19. Sparks, A.B., et al., *Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2.* Proc Natl Acad Sci U S A, 1996. **93**(4): p. 1540–1544.

20. Ren, R., Z.-S. Ye, and D. Baltimore, *Abl protein-tyrosine kinase selects the Crk adapter as a substrate using SH3-binding sites.* Genes and development, 1994. **8**(7): p. 783–795.

21. Rozakis-Adcock, M., et al., *The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1*. Nature, 1993. **363**(6424): p. 83-85.

22. Gout, I., et al., *The GTPase dynamin binds to and is activated by a subset of SH3 domains.* Cell, 1993. **75**(1): p. 25-36.

23. Pisabarro, M.T., L. Serrano, and M. Wilmanns, *Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions.* Journal of Molecular Biology, 1998. **281**(3): p. 513-521.

24. Feng, S., et al., *Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions.* Science, 1994. **266**(5188): p. 1241-1247.

25. Rickles, R.J., et al., *Phage display selection of ligand residues important for Src homology 3 domain binding specificity.* Proc Natl Acad Sci U S A., 1995. **92**(24): p. 10909-10913.

26. Wu, X., B. Knudsen, and S. Feller, *Structural basis for the specific interaction of lysine containing proline-rich peptides with the N-terminal SH3 domain of c-Crk.* Structure, 1995. **3**(2): p. 215-226.

27. Wu, Y., S.D. Spencer, and L.A. Lasky, *Tyrosine Phosphorylation Regulates the SH3-mediated Binding of the Wiskott-Aldrich Syndrome Protein to PSTPIP, a Cytoskeletal-associated Protein.* J Biol Chem, 1998. **273**(10): p. 5765-5770.

28. Nguyen, J.T., et al., *Exploiting the Basis of Proline Recognition by SH3 and WW Domains: Design of N-Substituted Inhibitors.* Science, 1998. **282**(5396): p. 2088-2092.

29. Mongioví, A.M., et al., *A novel peptide-SH3 interaction.* EMBO, 1999. **18**(19): p. 5300–5309.

30. Houdusse, A., A. Bahloul, and E. Ostap, *The crystal structure of the sh3 domain of acanthamoeba myosin ib bound to acan125.* Unpublished work, 2007.

31. Ferguson, M.R., et al., *Directed discovery of bivalent peptide ligands to an SH3 domain.* Protein Science, 2004. **13**(3): p. 626–632.

32. Tolia, N.H. and L. Joshua-Tor, *Strategies for protein coexpression in Escherichia coli.* Nat Meth, 2006. **3**(1): p. 55-64.

33. Feng, W. and M. Zhang, *Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density.* Nat Rev Neurosci, 2009. **10**(2): p. 87-99.

34. Pantoliano, M.W., et al., *High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery.* Journal of Biomolecular Screening, 2001. **6**(6): p. 429-440.

35. Lo, M.-C., et al., *Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery.* Analytical Biochemistry, 2004. **332**(1): p. 153-159.

36. Cimmperman, P., et al., *A Quantitative Model of Thermal Stabilization and Destabilization of Proteins by Ligands.* 2008. **95**(7): p. 3222-3231.

37. Jackson, M. and H.H. Mantsch, *Beware of proteins in DMSO.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1991. **1078**(2): p. 231-235.

38. Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases.* Biopolymers, 2008. **89**(5): p. 392-400.

39. Tjernberg, A., et al., *DMSO-Related Effects in Protein Characterization.* Journal of Biomolecular Screening, 2006. **11**(2): p. 131-137.

40. Godreau, D., et al., *Different Isoforms of Synapse-associated Protein, SAP97, Are Expressed in the Heart and Have Distinct Effects on the Voltage-gated K+ Channel Kv1.5*, 2003. p. 47046-47052.

41. Senthilkumar, R. and K.K. Sharma, *Effect of Chaotropic Agents on the Structure-Function of Recombinant Acylpeptide Hydrolase.* Journal of Protein Chemistry, 2002. **21**(5): p. 323-332.

42. Kraft, C.A., et al., *Quantitative Analysis of Protein-Lipid Interactions Using Tryptophan Fluorescence.* Sci. Signal., 2009. **2**(99): p. pl4-.

43. Garcia, E.P., et al., *SAP90 binds and clusters kainate receptors causing incomplete desensitization.* Neuron, 1998. **21**(4): p. 727-39.

44. Bartos, J.A., et al., *Postsynaptic Clustering and Activation of Pyk2 by PSD-95.* The Journal of Neuroscience, 2010. **30**(2): p. 449-463.

45. Sun, Y., et al., *Polyglutamine-expanded Huntingtin Promotes Sensitization of N-Methyl-d-aspartate Receptors via Post-synaptic Density 95*, 2001. p. 24713-24718.

46. Shin, H., et al., *An Intramolecular Interaction between Src Homology 3 Domain and Guanylate Kinase-Like Domain Required for Channel Clustering by Postsynaptic Density-95/SAP90.* The Journal of Neuroscience, 2000. **20**(10): p. 3580-3587.

47. Garcia, E.P., et al., *SAP90 Binds and Clusters Kainate Receptors Causing Incomplete Desensitization.* Neuron, 1998. **21**(4): p. 727-739.

Chapter 8. Biochemical and structural characterisation of B Lymphocyte Stimulating Factor (BLyS)

8.1 Introduction

BLyS (also known as BAFF, THANK, TALL-1,zTNF4 and TNF13B) is a protein ligand that belongs to the tumour necrosis family (TNF) which plays a functional role in B cell immunity and has recently come under study as a target for treatment of autoimmune disease, most notably systemic lupus[1, 2].

Action of BLyS is mediated through binding to the relevant TNF receptor. BLyS is known to bind to 3 receptors; BAFF-R, TACI (transmembrane activator and calcium modulator ligand interactor) and BCMA (B cell maturation antigen), however binding to TACI and BCMA requires multimerisation of the BLyS trimer.

As a consequence of BAFF receptor activation through BLyS, B cells receive a primary survival signal. Mice which overexpress BLyS have increased numbers of B cells produced in many tissues[3]. Consequently, if the regulation of BLyS production becomes compromised and upregulation of BLyS takes place, the survival signals triggered by BLyS can override death signals to autoreactive B cells giving rise to autoimmune disease[4].

Therefore, if the binding interface between BLyS and BAFF-R is targeted for disruption by small molecules, there is potential for the development of orally active drugs for treatment of autoimmune disease that would be more cost efficient and convenient to use than antibody treatments that are currently in development.

8.2 Druggability of the BLyS/BAFF-R complex

The crystal structure of BLyS in complex with BAFF receptor fragments has been solved (PDB ID 1POT)[5]. From the crystal structure it is known that BLyS and the BAFF receptor bind in a 1:1 stoichiometry, furthermore the interaction between both proteins is not affected by multimerisation. This is an important feature as BLyS can exist in monomeric, trimeric and 60-mer states[6]. Interaction between BLyS and the BAFF receptor is mediated through conserved cysteine rich domains (CRDs) from the receptor with the TNF homology domain (THD) from BLyS[7]. The CRDs usually contain in the region of six cysteine residues which form three disulphide bonds and elongate the shape of the receptor, the positions of the cysteines are not conserved. In the BAFF receptor four cysteines are found in the CRD which form two disulphide bonds (Figure 8-1).

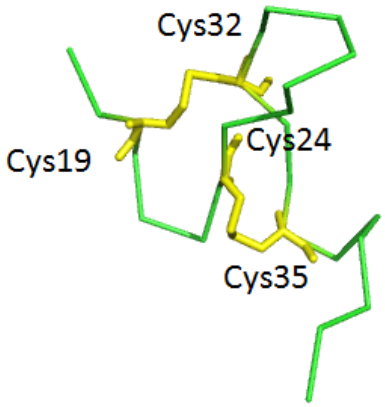


Figure 8-1. The cysteine rich domain (CRD) of the BAFF receptor. This creates a hairpin structure containing the DxL motif which is an essential feature of BLyS/BAFF-R binding[8].

The TNF homology domain of BLyS (Figure 8-2) is composed of ten β-strands which are arranged as two antiparallel β-sheets with a greek key topology[9]. Strands A, A′, H, C and F form the innermost sheet which plays a role in establishing trimer

contacts, while strands B, B′, D, E and G are surface exposed and make up the outer layer of the sheet. The DE-loop of BLyS is more extended than in other THDs, this allows the protein to form multimers composed of many trimers.

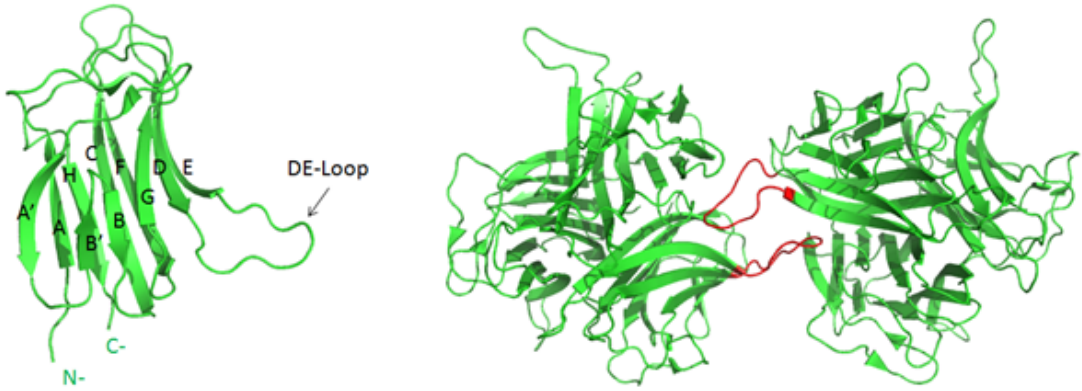


Figure 8-2. The TNF homology domain of the BLyS monomer (left). The extended DE loop allows interactions between two BLyS trimers in a "handshaking" fashion.

If the BLyS trimer structure is examined (Figure 8-3) it can be noted that the BAFF receptor binds to each monomer via a series of hydrogen bonds which are mediated by Asp26 and Leu28 of BAFF-R and Arg231, Arg265 and Tyr206 of BLyS.

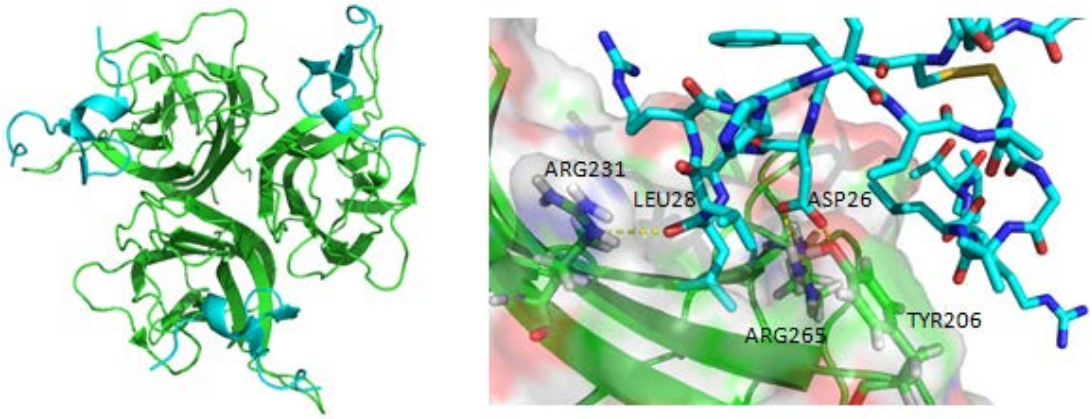


Figure 8-3. Structure of BLyS trimer complexed with BAFF-R fragments. The BLyS (coloured green) binds to the BAFF-R fragment (coloured cyan) through a network of hydrogen bonds that involve Arg231, Arg265 and Tyr206 of the BLyS protein and Asp26 and Leu28 of BAFF-R.

8.3 Expression, purification and characterisation of BLyS

8.3.1. Expression

BL21 (DE3), BL21 Star and Rosetta Gami B cells were transformed with plasmid encoding the full length BLyS protein with an N-terminal histidine tag. Transformed colonies were selected for using 50µg/ml kanamycin. Overexpression optimisation was performed using three cell lines, four types of growth media, two IPTG concentrations and three temperatures as shown in Figure 8-4. Cells were grown until OD600 reached approximately 0.8, and then expressed for 17 hours.

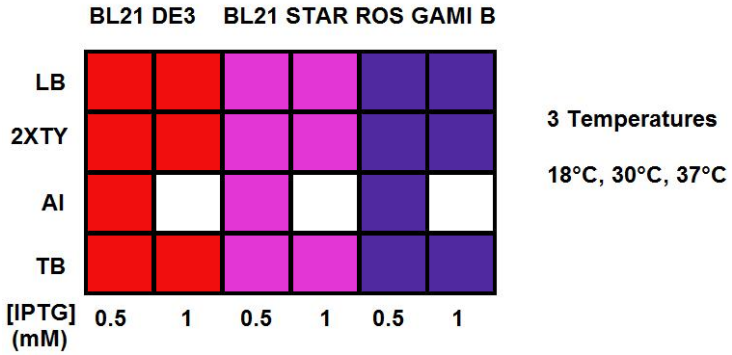


Figure 8-4. Overexpression trials of BLyS protein. In total 63 conditions for overexpression were evaluated in this trial. Growth media used included standard LB, 2XTY, Autoinduction media (Novagen) and Terrific Broth (TB). Autoinduction media does not require IPTG to induce protein expression and therefore only one culture was set up for each cell line in this media.

The soluble BLyS fraction was still very low yield (< 0.3 mg/ L), however the difference in insoluble protein between the lowest and highest yield was threefold. The top three conditions for BLyS production are shown in Figure 8-5

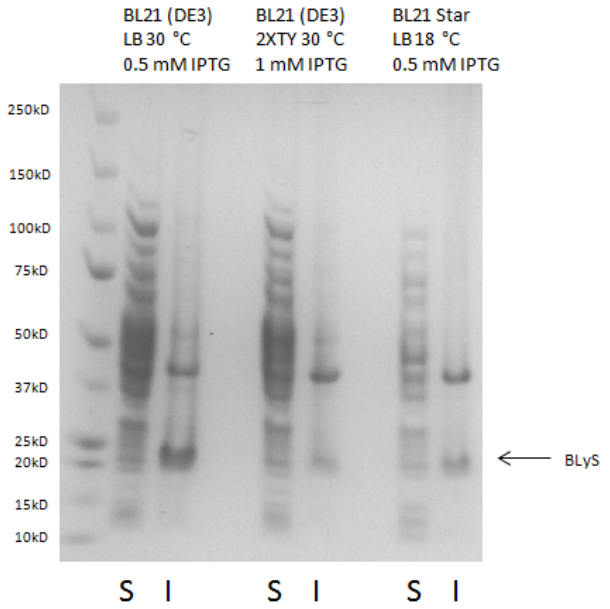


Figure 8-5. Overexpression trials for BLyS protein showed protein is still insoluble. The gel shows the Soluble (S) and Insoluble (I) fraction for each of the expression conditions listed. After evaluating 64 conditions there was minimal increase in soluble BLyS protein.

8.3.2 Refolding and purification

The majority of the BLyS protein was insoluble so refolding the protein was required. Insoluble cell pellets were prepared as follows: Whole cell pellets were resuspended in 50mM $NaPO_4$ pH 7.4 and lysed by cell disruption at 22KPSI. Lysate was centrifuged at 30,000 xg for 20 minutes and the insoluble pellet retained and solubilised in either 8M urea or 6M GnHCl depending on method used. Three refolding protocols were tested to compare their effectiveness; the first was an approach based on folding by dilution whilst the second and third used on-column refolding.

Protein refolding by dilution

Insoluble BLyS inclusion bodies were resuspended in 8M urea at 4 °C with stirring for 1 hour. The resuspended pellet was then filtered sequentially through 0.45 µm

and 0.22 µm filters and then loaded dropwise into refolding buffer (50 mM HEPES, 10% glycerol, 500mM NaCl, 0.25mM NDSB, pH 8) at a 1:10 dilution.

On column refolding

Two types of on column refolding were compared. In the first instance insoluble BLyS was resolubilised in 8M urea, filtered through 0.45 µm and 0.22 µm filters before loading on an IMAC HiTrap FF column charged with nickel sulphate. BLyS protein was refolded using a gradient of 6-0 M urea in refolding buffer (50 mM Tris, 5 mM EDTA, 150 mM NaCl, 1.25 mM GSH and 0.25 mM GSSG, pH 8). Protein was eluted from the column in a gradient from 50mM-500mM imidazole in PBS.

In the second instance insoluble BLyS was resolubilised in 6M GdnHcl, 10 mM DTT, centrifuged at 12,000 xg for 20 minutes and then filtered through 0.45 µm and 0.22 µm filters before loading on an IMAC HiTrap FF column charged with nickel sulphate. The refolding buffer was a gradient of 4-2M urea, 20 mM β-mercaptoethanol, 100 mM Tris, pH 8. Elution was performed with a gradient of 20-500 mM imidazole in 50 mM Tris, 150 mM NaCl, pH 8.

The purification itself was a two stage process which involved refolding the urea denatured protein and then eluting using a standard imidazole gradient as shown in Figure 8-6. The requirement for refolding the BLyS protein is documented in the literature with a number of refolding protocols being published[10, 11]. After trying both types of protocol as well as refolding by dilution[12] there was no significant difference between the success of each method as verified by gel filtration study. The volume of sample generated from refolding by dilution was however a disadvantage

to the technique as well as the inability to automate the process of refolding and purification like the on-column refolding protocol with the AKTA purifier.

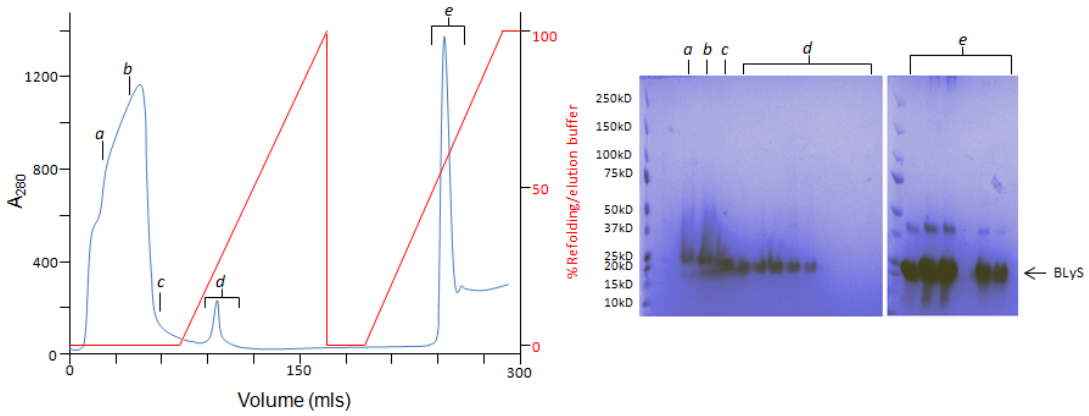


Figure 8-6. Purification results for on column refolding of BLyS from inclusion bodies. During the renaturation process some BLyS is lost from the column, however the final yield of refolded BLyS was 23 mg/ L culture. The first gradient shows the refolding step whilst the second is the imidazole elution gradient.

During the elution it was noted that a second protein purified alongside BLyS with a molecular weight of approximately 40kDa. A similar band is also observed in other refolding studies where *E.coli* is used as the expression system.

8.3.3 Characterisation of the BLyS 60-mer complex

8.3.3.1 Analytical gel filtration

500 µl BLyS sample was loaded onto a Superdex-75 10/300 column equilibrated with 50mM sodium phosphate, 150mM NaCl pH 8, and eluted at 0.5ml/ min.

Gel filtration analysis was used on each type of refolding experiment to assess possible differences in folding state that could occur between the methods. A side by side comparison of all UV traces are shown in Figure 8-7.

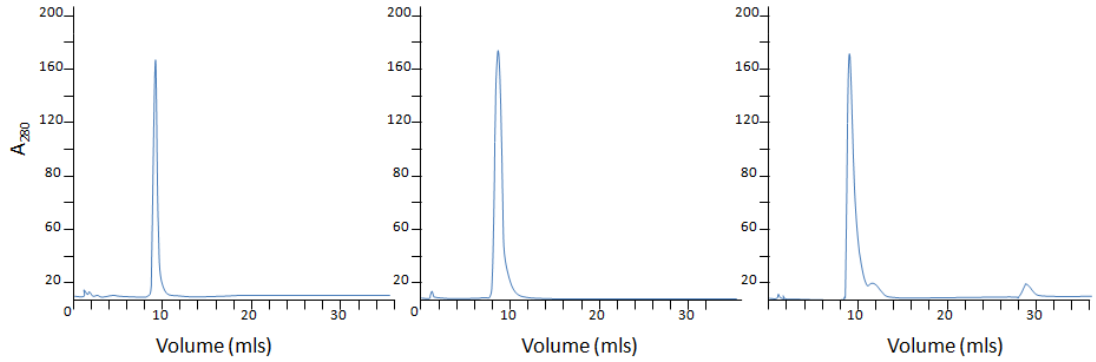


Figure 8-7. Comparison of refolding state of BLyS by gel filtration analysis. The on column refolding protocols produced almost identical elution profiles. From left to right: On column refolding using GSH GSSG, On column refolding using β-mercaptoethanol and 100 mM Tris. Finally refolding by dilution using HEPES buffer.

Elution of BLyS began at 8 ml which in the Superdex 75 10/300 corresponds to the void volume of the column where macromolecules larger than 600 kDa are excluded from the matrix. As BLyS is known to form a 60-mer it seemed likely that this was the state of the protein during the gel filtration process, the molecular weight of the 60-mer would be approximately 1.2 mDa. Alternatively it was also possible that the protein was being voided from the column as a result of aggregation.

Whilst there are some smaller peaks in the dilution method profile these were very low concentration and were not detectable in coomassie stained gels. To check the purity of the BLyS peaks from gel filtration samples were run on SDS-PAGE gel and silver stained (Figure 8-8). The largest bands in the refolded protein lanes were a similar size to the soluble BLyS, the refolding by dilution sampled contained two contaminants which had not been visible through coomassie staining, one protein of 50 kDa and the other approximately 60 kDa.

After gel filtration there was little contamination in the column refolded BLyS therefore it was suitable for further characterisation studies.

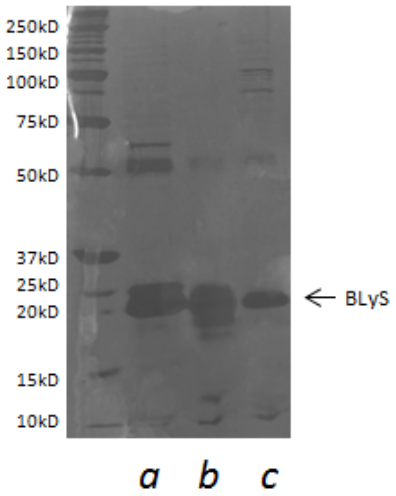


Figure 8-8. Silver stain results of refolded BLyS in comparison to soluble BLyS. (*a*) Refolded by dilution. (*b*) On column refolding (*c*) BLyS protein in soluble fraction.

8.3.3.2 Thermal denaturation

6.5 µM BLyS protein was used in each sample well with 50mM sodium phosphate, 150mM NaCl pH 8 buffer. Melting profiles were obtained in triplicate for samples which had been incubated with 10X Sypro orange dye for one hour at 4 °C and for samples which had been incubated with the dye for 90 seconds prior to loading into the thermocycler. The effect of pH was also measured as a variable in the range of 6.4-7.4.

In order to assess whether the protein was in the form of a 60-mer or in an aggregated state a thermal melt was performed. The melt results were not what would typically be expected (Figure 8-9). Only at pH 6.4 and pH 6.8 with 1 hour incubation is there a peak which may be attributed to a melting event at 60 °C. Initially it could be assumed that this is because the protein is aggregated, however if the structure of the 60-mer is examined more closely (Figure 8-10) it is understandable why the results obtained are atypical.

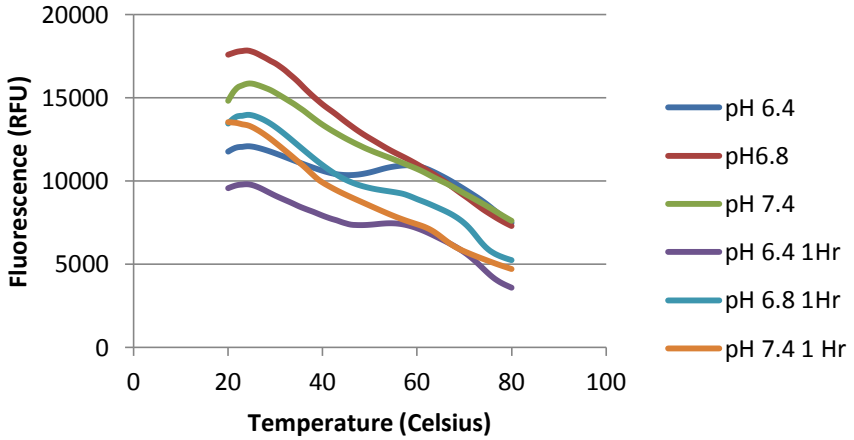


Figure 8-9. Melting profile of refolded BLyS protein with varying pH. Two sets of data were recorded for each pH value. The first was from the protein sample being mixed with Sypro orange and then immediately loaded into the thermocycler, in the second Sypro was incubated with the protein for one hour prior to the start.

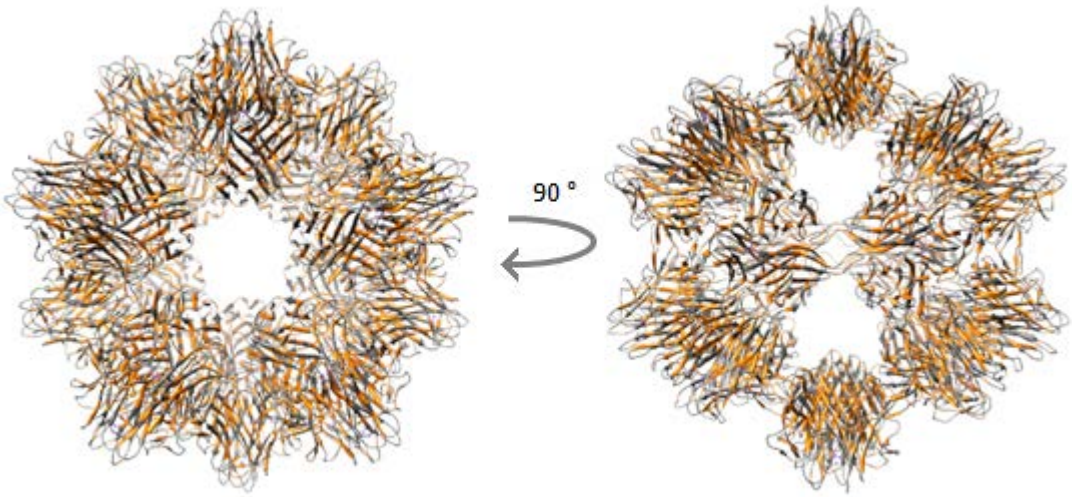


Figure 8-10. Hydrophobic profile of BLyS 60-mer. Taking the PDB structure for the 60-mer 10TZ and colouring the hydrophobic residues orange, it is apparent that the virus like macromolecule presents significant hydrophobic surface area to solution.

The BLyS 60-mer structure does not just present a number of hydrophobic surfaces on the outer structure, there are also 11 channels into the protein with a minimum diameter of 31.6 Å which make the inner space very solvent accessible.

Both the large area of hydrophobic contacts on the outer surface as well as a similar surface area on the inner surface create an environment that would make the Sypro orange dye fluoresce. Therefore the results in Figure 6-6 could be explained not just by presence of aggregated protein but also by Sypro orange binding to the numerous hydrophobic sites present on the correctly folded 60-mer complex. The difference in behaviour observed as a consequence of increased incubation time may be explained by the preparation of the one hour incubations at 4 °C rather than room temperature, which would suggest that the BLyS 60-mer is unstable at room temperature.

8.3.3.3 Intrinsic fluorescence spectroscopy

Each BLyS monomer contains one tryptophan residue, the residue is solvent exposed but not within the target binding site. Changes in fluorescence observed in the presence of compounds cannot give a positive indication that the compound is binding in the active site, however it gives a general indication that there may be a conformational change occurring.

8.3.3.4 Comparisons of emission spectra between refolded and denatured BLyS

Fluorescence from PBS buffer was measured and then subtracted fom the spectra for 2.6 µM BLyS in PBS buffer. Similarly 8M urea in PBS was subtracted from the 2.6 µM BLyS in 8M urea sample.

To gain a better understanding of the difference between a denatured (and therefore incorrectly folded) BLyS sample and the sample from refolding experiments the intrinsic fluorescence value of BLyS was measured in a denatured and refolded state (Figure 8-11a).

The fluorescence profile for the refolded BLyS was consistent with a maxima of approximately 1.4 million RFU at 336.5 nm. The denatured BLyS fluorescence was quite different. The maxima position had not only been red shifted from 336.5 nm to 350 nm but there was also a 14% decrease in fluorescence overall. Emission at longer wavelength is characteristic of the tryptophan residue becoming more solvent exposed[13] which would agree with the protein becoming more disordered as a result of the chaotropic agent. From this result a clear difference between the refolded and unfolded states can be seen.

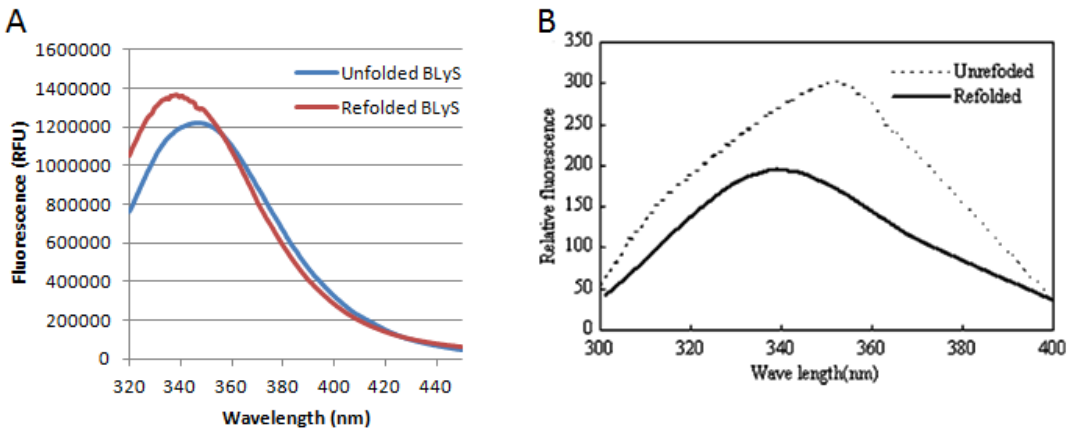


Figure 8-11. Fluorescence profiles of urea denatured and refolded BLyS. (a) Results obtained from this study. (b) Results obtained by Cao et al.

Fluorescence results from this study do not wholly correlate with those obtained from Cao et al[10]. Whilst unfolding the protein causes a red shift in both cases, Cao et al observed an increase in fluorescence which is in contrast to that reported here. Despite the authors of the other study using a protein concentration which was twice that reported here, the UV trace observed was of low intensity and the curve uncharacteristic of a typical tryptophan fluorescence measurement. While the use of an excitation wavelength of 295 nm was used in this study to minimise fluorescence

effects from the tyrosine and phenylalanine residues the recorded fluorescence from tryptophan should have still overwhelmed any signal from these residues at the 280 nm excitation wavelength used by Cao et al. Closer inspection of the spectra (Figure 8-10 b) does show that there is no discernable peak at 303 nm which would correlate with tyrosine emission and the starting wavelength of measurement is higher than the emission peak of phenylalanine at 282 nm. Given the low intensity of the signal there is the possibility that a low signal to noise ratio is playing a factor.

8.3.3.5 Transmission electron microscopy (TEM) analysis

In order to assess the protein concentration which would give best resolution of BLyS multimeric state with the least aggregation protein concentrations of 216 µg / ml, 635 µg/ ml, 1.350 mg/ ml and 15 mg/ml were used. Protein samples were applied to copper grids which were previously glow discharged and carbon coated. All samples were visualised using negative staining with 2 % uranyl acetate on a Phillips F20 Biotwin transmission electron microscope.

As the BLyS 60-mer is such a large macromolecule it can be directly visualised by electron microscopy. By preparing a selection of BLyS concentrations for TEM the concentration which contained a representative population of the protein state could be found. The 216 µg / ml and 635 µg/ ml samples (Figure 8-12 *a* and *b*) were too low concentration to accurately assess if the protein had significant aggregation therefore concentrations at 1.350 mg/ ml and 15 mg/ ml were also used (Figure 8-12 *c* and *d*). These higher concentrations allow some aggregation to be seen as darker grey regions on the images, however even at 15 mg/ml concentrations the aggregated protein makes up only a small proportion of the protein sample.

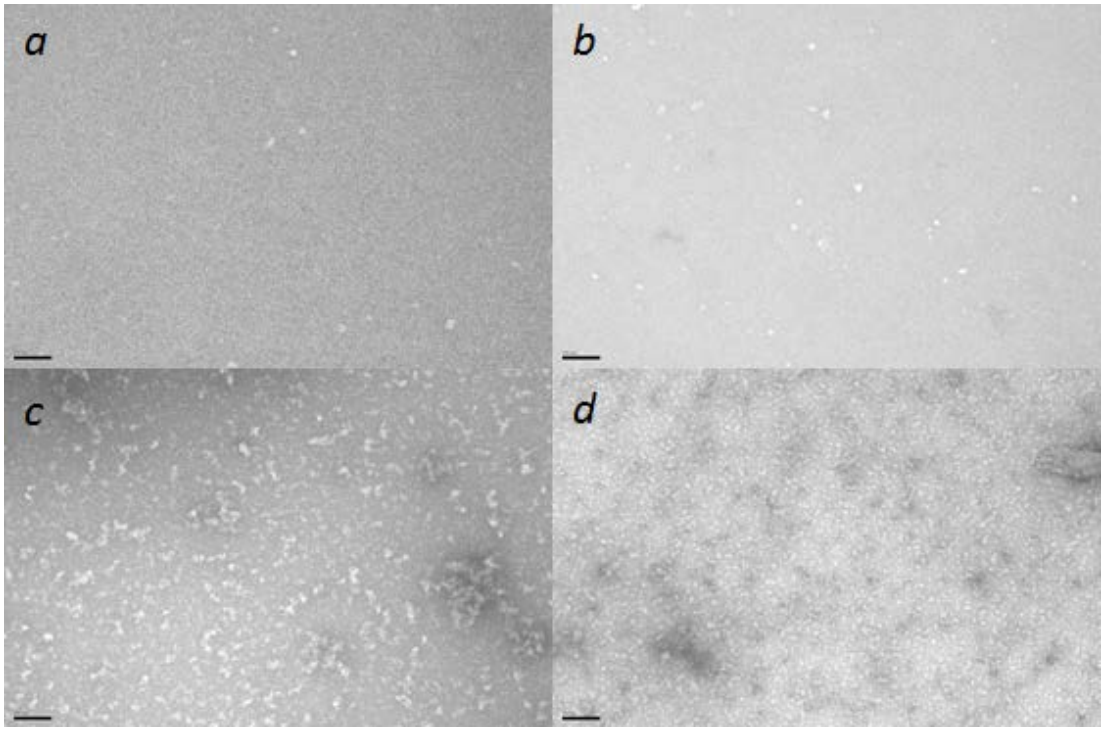


Figure 8-12. Transmission electron microscope images of BLyS concentration range. (*a*) BLyS concentration 216 µg / ml. (*b*) BLyS concentration 635 µg/ ml. (*c*) BLyS concentration 1.35 mg/ ml. (*d*) BLyS concentration 15 mg/ml. Scale bar shown is 100 nm.

When a high resolution image of the BLyS sample is compared to that from the literature there is a good correlation between the observed shapes from this study and the published structure which measured approximately 20nm in diameter[14].

8.3.3.6 Dynamic light scattering

BLyS protein was buffer exchanged into PBS (0.14 M NaCI, 0.01 M PO_4 Buffer, 0.003 M KCI) and concentrated to 0.3mg/ ml. The sample was then clarified by centrifugation at 13,000 RPM in a chilled benchtop centrifuge. Measurements were recorded on a Malvern Zetasizer APS using 100 µl sample volumes in a 96 well plate format.

Two peaks were visible from the DLS results (Figure 8-13) which have an apparent molecular weight of approximately 1.2 mDa and 73 mDa respectively according to the Mark-Houwink distribution equation. The higher molecular weight peak is likely some of the aggregated protein, as seen in the TEM results, which has reformed after gel filtration. Analysis of the sample by volume shows that the aggregated protein makes up very little of the sample as it appears as a single peak with an approximate diameter close to 20 nm which is in agreement with the measured diameter of the crystal structure of the BLyS 60-mer complex.

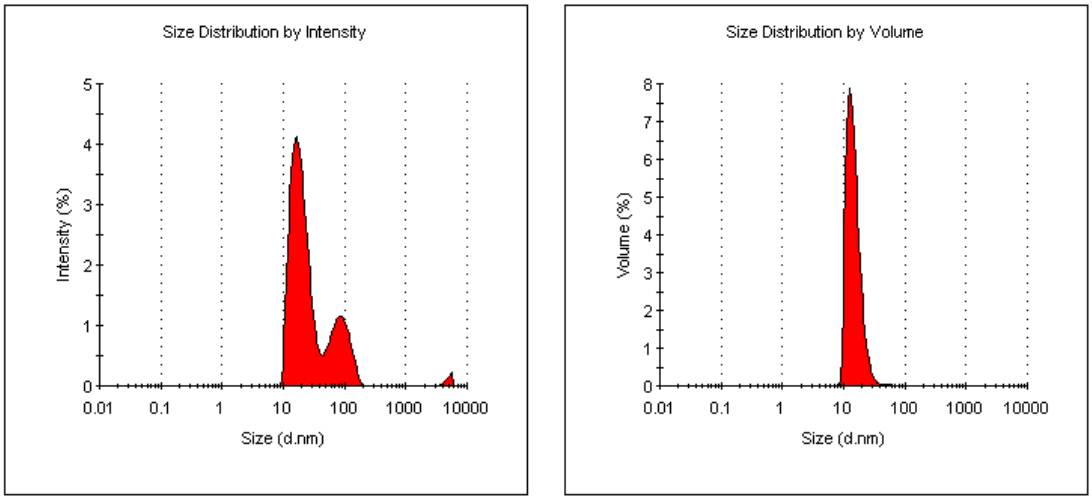


Figure 8-13. Dynamic Light Scattering (DLS) results show two main peaks in solution.

8.4 Virtual screening of the BLyS structure with LIDAEUS

8.4.1 Defining the site points for BLyS

Using the crystal structure of the BLyS/BAFF-R complex (PDB codes 1OTZ and 1POT) the site points for LIDAEUS docking were created as described in Chapter 2 (Figure 8-14). A particular point of interest is the pocket which is located at the position where Leu28 interacts with Arg231. The depth of the binding pocket is a key concern when undertaking small molecule screening and drug design as the shallowness of the pocket determines how "druggable" the site is[15].

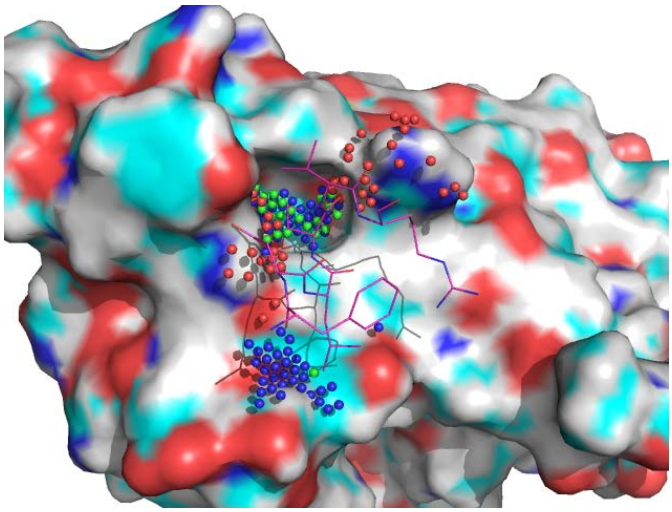


Figure 8-14. Defining site points on the BLyS/BAFF-R interface. The BAFF-R fragment (sequence FDLLVR) used to define the site points which will guide LIDAEUS docking is shown in line format coloured purple. The pocket in the surface at the Leu28 position is a particularly attractive target for small molecule interaction. Green spheres represent hydrophobic sites, blue represents hydrogen bond donors and red represents hydrogen bond acceptors.

8.4.2 Analysis of the Leu28 binding pocket reveals a suitable target for docking

A number of programs have been developed to evaluate the depth of a protein surface pocket for suitability in small molecule ligand binding[16-18]. These programs

allow preliminary assessment of a protein's drugability as a function of the characteristics of the pockets found on the protein surface (depth, solvent accessibility, etc). The surface of BLyS was evaluated for binding pockets using the PocketDepth program which found seven possible pocket regions (Figure 8-15) including the Leu28 binding region highlighted previously which shall be referred to as pocket 5.

Whilst the six other locations would not be suitable targets for disruption of the BLyS/BAFF interaction because they are located too far away from the binding site, there is a possibility that they could be targeted for disruption of the BLyS 60-mer which could be a useful way of interrupting the association of the BLyS multimer with TACI.

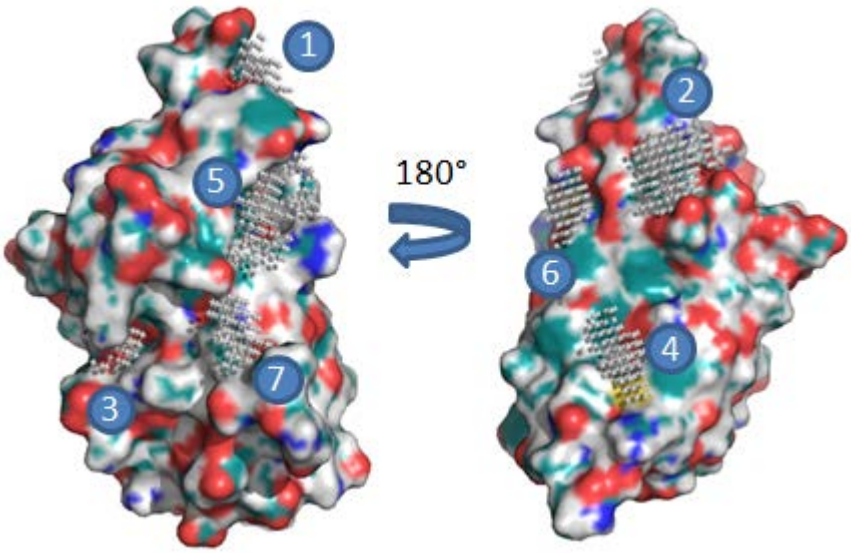
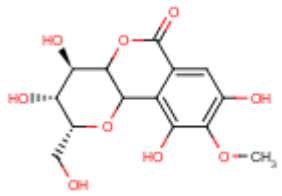
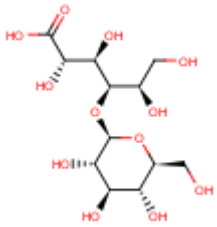


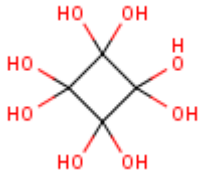
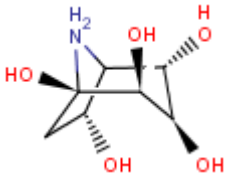
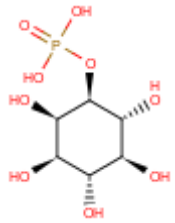
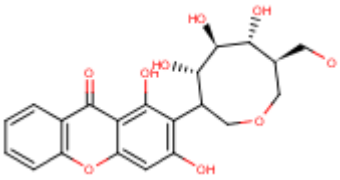
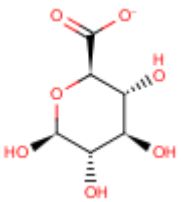
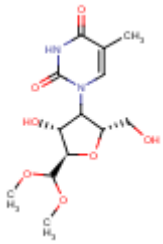
Figure 8-15. Binding pocket predictions from PocketDepth analysis on BLyS. Seven locations were highlighted by the program one of which is the Leu28 binding pocket (Position 5).

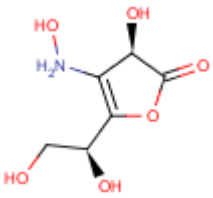
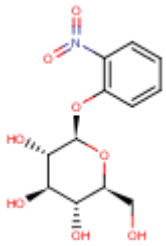
8.4.3 LIDAEUS docking of the EDULISS database to the BLyS protein

The LIDAEUS run against the entire EDULISS database was performed using >4m small molecule compounds as described in Chapter 2. The top 200 compounds were kept and assessed by eye how well they fit into the Leu28 binding pocket. The most interesting compounds from this are shown in Table 8-1. The compounds selected by LIDAEUS were then scored using Autodock. The Autodock results were slightly different from those obtained from LIDAEUS, often the highest scoring pose was slightly different from that suggested by LIDAEUS (although still located within the target pocket).

Table 8-1. Structures and scores for the top 10 scoring compounds from BLyS/BAFF-R binding pocket docking.

Structure	LIDAEUS score (kcal /mol)	Autodock score	Compound ID
	-53.9	-4.6	25SPH1-178-375
	-49.4	-4.1	26SPH1-825-056

	-48.9	-4.5	25SPH1-115-197
	-47.5	-4.4	35SPH1-122-641
	-46.6	-4.9	25SPH1-007-145
	-46.4	-5.4	26SPH1-677-512
	-44.8	-4.6	25SPH1-061-217
	-43.1	-5.1	35SPH1-131-341

	-42.6	-4.3	35SPH1-032-786
	-41.8	-5.0	25SPH1-072-060

Of the 200 compounds retained from the LIDAEUS docking 25SPH1-115-197 (Cyclobutaneoctol) appeared 18 times. The small ring structure is an ideal shape for binding into the target pocket, however it is also a compound commonly used as a pesticide and therefore its use as a drug is limited. Looking at compounds 25SPH1-178-375, 26SPH1-825-056 and 26SPH1-825-056 in more detail the hydrogen bond networks can be observed (Figure 8-16 to 8-18).

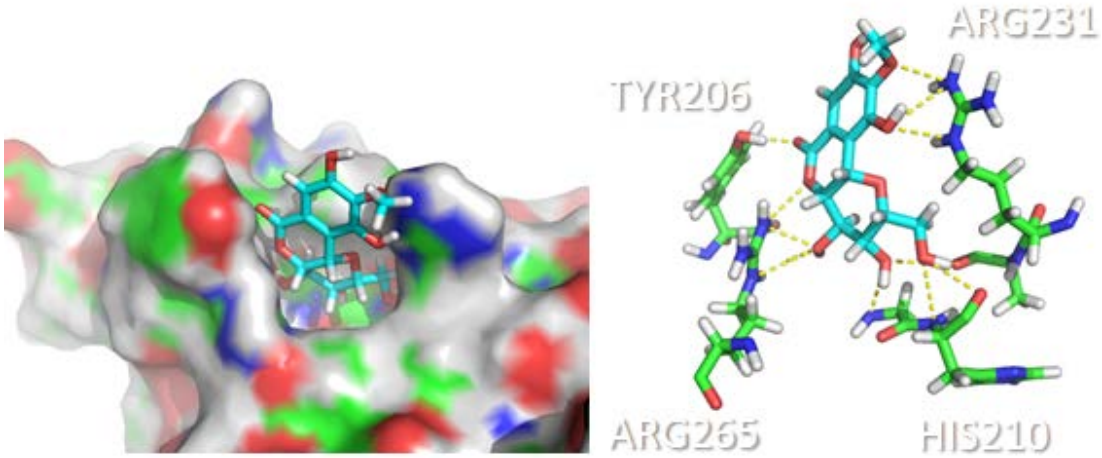


Figure 8-16. Predicted binding mode between 25SPH1-178-375 and BLyS shows a number of hydrogen bonding opportunities both deep inside the pocket as well as the upper lip.

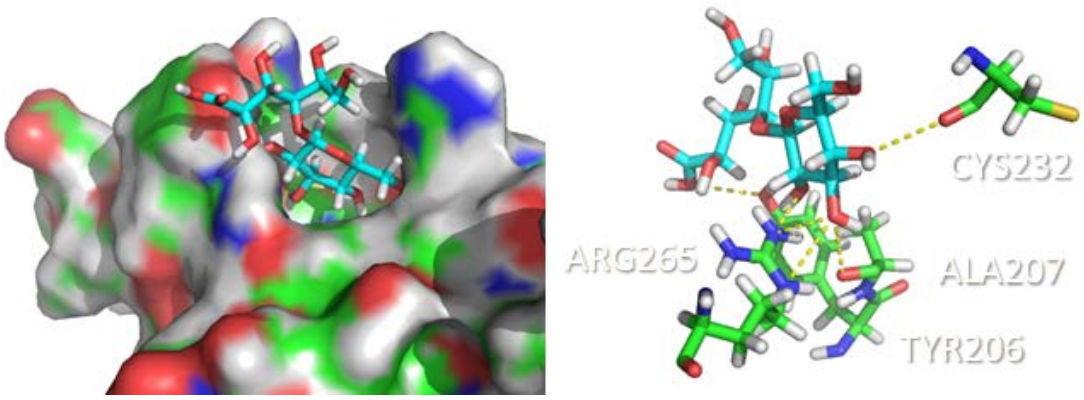


Figure 8-17. Predicted binding mode between 26SPH1-825-056 and BLyS shows a number of hydrogen bonding opportunities both deep inside the pocket as well as the upper lip.

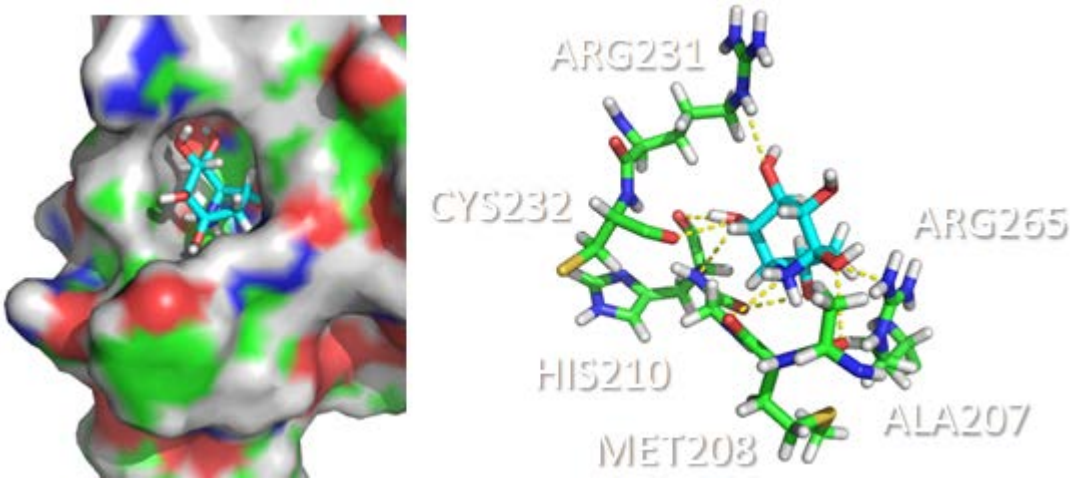


Figure 8-18. Predicted binding mode between 35SPH1-122-641 and the BLyS protein. 8 Hydrogen bonds are possible between the compound and the protein.

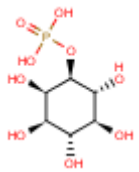
The LIDAEUS docking results indicate that a ring structure is an important part of anchoring the small molecule into the pocket. The ring alone is able to access the pocket fairly well, however it is also likely to be able to enter one of the other 6 pockets on the BLyS surface. To test this, structures for 35SPH1-122-641 and 25SPH1-072-060 were docked into the pockets from Figure 8-15 and were found to have similar affinity scores for some of the other pockets on the protein. As such

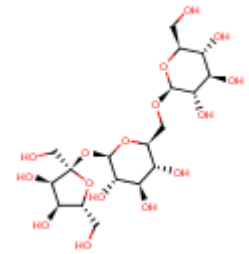
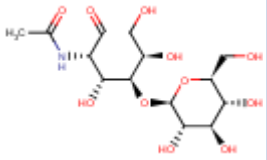
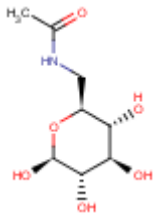
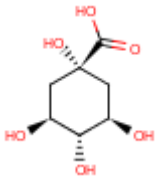
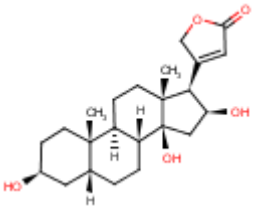
compounds 25SPH1-178-375, 26SPH1-825-056 and 26SPH1-677-512 are more likely to be ligands with better specificity for the target binding site than the other compounds in Table 8-1

8.5 Ligand selection for BLyS/BAFF-R disruption screening

Only a small selection of the top 10 compounds from Table 8-1 were readily available for in vitro testing. All compounds for this study were purchased from Sigma-Aldrich and were available for immediate dispatch. The compounds available for in vitro screening are shown in Table 8-2 and accounted for a good spectrum of the top 100 results. The final compound in the table was included as a compound of interest from a similar study conducted by Wissam Mehio[19] which was not able to be tested previously but was a high scoring molecule in that docking study.

Table 8-2. List of compounds selected for assay with BLyS protein.

Structure	EDULISS ID	Sigma No	LIDAEUS Rank
	SPH1-007-145	F8502	13

	SPH1-171-521	R0250	16
	SPH1-100-780	A6259	43
	SPH1-100-826	A6645	59
	SPH1-168-399	138622	71
	SPH1-104-152	G4635	-

Hexopyranose sugar moieties are not generally regarded as an optimal synthetic starting point in drug discovery, usually due to their high polarity and poor pharmokinetics. However there has been interest in their use as a scaffold[20] which

can provide a source of molecular diversity[21]. Glycomimetic drugs which have better bioavailability and improved plasma half life over traditional carbohydrates have been developed. These drugs are able to exploit the recognition of specific carbohydrate sequences on cell surfaces in their mechanism of action, this led to the development of the antiviral drug Tamiflu[22]. Compounds with a sugar-like moiety have found to disrupt protein-protein interactions in the SH3 domain of PSD95[23]. Recent advances in the synthesis of saccharides shows potential in the development of novel antiviral, antibacterial and anticancer therapeutics[24].

8.6 Measuring the effect of virtual screening compounds on BLyS fluorescence

Fluorescence measurements were taken using a Fluoromax-3 specrofluorometer (Horiba Jobin Yvon) with a 5nm slit width. Total sample volume was 600 µl in a 1 cm path length quartz cuvette at 10 °C. The excitation wavelength was 295 nm and the emission range was from 320-450 nm in 0.5 nm increments. Readings were taken in triplicate and then averaged.

When measuring compound fluorescence the six compounds obtained from virtual screening were dissolved at a concentration of 20 mM in 100% DMSO, except for compound F-8502 which was not fully soluble in DMSO but soluble in 100% ddH_2O.

Fluorescence of all compounds at a concentration of 250 µM in PBS was measured by excitation at 295 and measuring emmision from 320-450nm to allow subtraction from the protein and compound mixture. A DMSO concentration of 1.25% was kept consistent accross all samples including the compound which was dissolved in water.

The effect on BLyS fluorescence by each compound was measured by mixing 2.6 µM BLyS with 250 µM of the test compound, briefly mixing by pippetting and incubating on ice for 10 minutes before starting measurement.

The fluorescence spectra obtained for the dissolved compounds (Figure 8-19) showed that all but two of the compounds had very little fluorescent activity over the Raman scatter from the buffer.

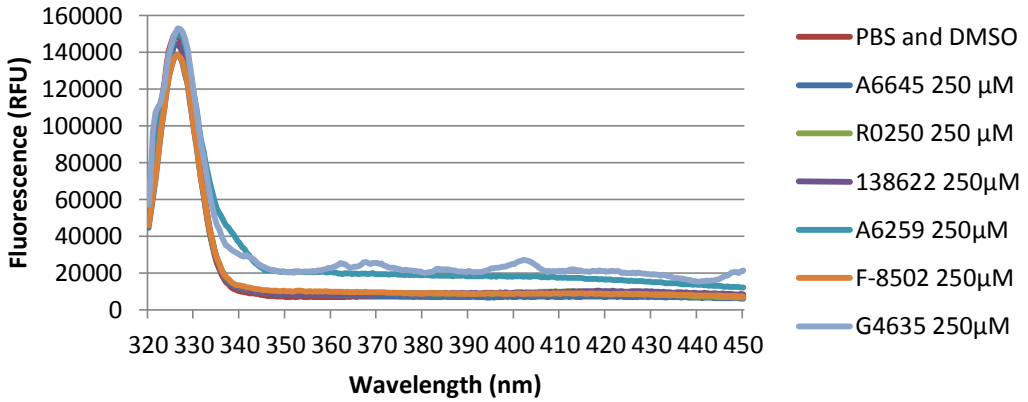


Figure 8-19. All six compounds chosen for BLyS fluorescence study have very little fluorescent properties. The advantage of this is that a significant change in fluorescence can be directly attributed to a change in the environment of the tryptophan residues in BLyS.

The compound effects on BLyS fluorescence were interesting as a number of different effects that were observed (Figure 8-20). Compound F8502 was the first to be tested and showed no change in fluorescence compared to the reference cell containing BLyS and 1.5% DMSO. This was also the case for compounds A6645 and 138622. Compound R0250 caused a much more noticeable change in fluorescence where a 20% decrease in signal was observed.

Conversely compounds A6259 and G4635 both caused a noticeable increase in fluorescence.

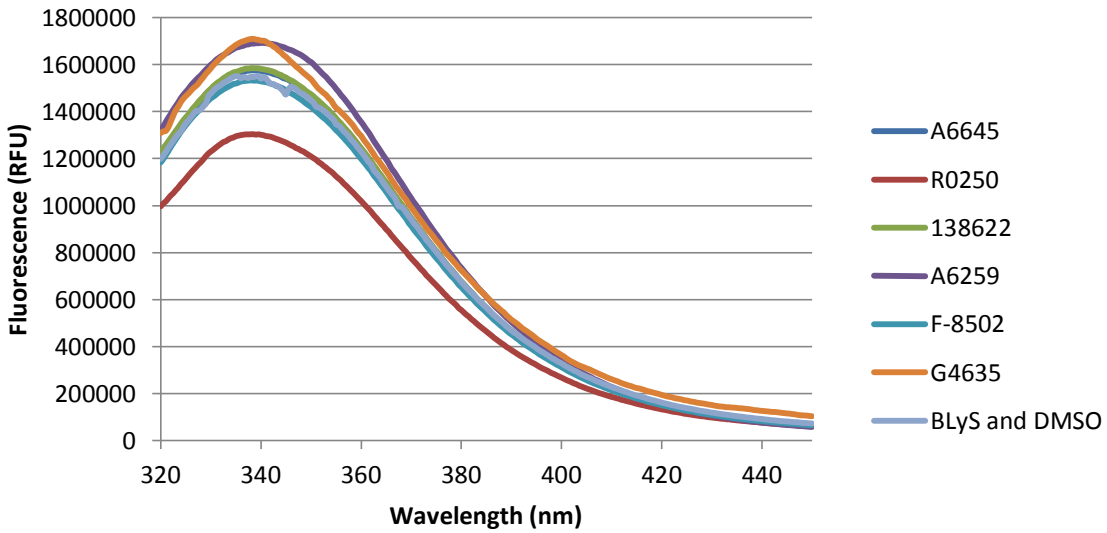


Figure 8-20. Effects of compounds from virtual screening on BLyS protein fluorescence.

Although no change in fluorescence was observed for compounds F8502, A6645 and 138622 this could be because of their highly hydrophilic nature which may cause competition with water for the binding site.

Compounds A6259 and G4635 caused an increase in fluorescence. These compounds have the potential to bind covalently to BLyS via the aldehyde functional group (compound A6259) and via a Michael addition to the C=C bond in the heterocyclic ring (compound G4635), although neither reaction has been tested. A number of drugs have been developed which feature the aldehyde functional group in the primary structure or in metabolic intermediates; the HIV inhibitor Abacavir[25] and the anticancer drugs Cyclophosphamide[26], Ifosamide[27] and Misonidazole[28] all feature this moiety. Outwith the scope of drug development such compounds can also be useful biochemical tools.

The other compounds appeared to have no activity with BLyS apart from compound R0250 which caused a 20% decrease in fluorescence. A large decrease in tryptophan

fluorescence is usually observed when the tryptophan residue becomes less solvent exposed, as a result of conformational changes in the protein or by direct ligand association with the tryptophan itself. In BLyS the tryptophan residue is located near pocket 3 (Figure 8-14) which is on the opposite side of the protein to the target binding site in pocket 5 (Figure 8-14). Docking the R0250 molecule into pocket 5 shows a low affinity of -3.5 kcal/ mol as the molecule is unable to enter the pocket itself. Association with the tryptophan on the outer surface of the pocket is possible but the affinity is much weaker compared to the target binding site.

8.7 Conclusions and future work

The BLyS protein is an interesting target not only from a disease treatment standpoint, but also because of the novel multimerisation that it can undergo. After refolding and purification the BLyS multimer was characterised by electron microscopy to confirm the multimer topology. Thermal denaturation gave further insight into the large hydrophobic area which is presented in the 60-mer compex which was found to readily bind Sypro orange dye. Dynamic light scattering also showed that the population of the sample was largely in this multimer state.

As a virtual screening target the BLyS monomer had a number of possible druggable pockets, however the pocket containing Leu28 was the focus of this study as it controls association with the BAFF receptor. A number of saccharide scaffolds were tested for binding activity and one of these caused a large decrease in fluorescence which could be attributed to ligand binding. Increases in fluorescence were also shown with two compounds which may form covalent bonds to BLyS.

The binding activity of these compounds could be further explored by using the CRD domain of the BAFF receptor, or a smaller peptide sequence that mimics the BAFF receptor CRD domain in a competition binding experiment. In addition the use of co-crystallisation techniques between BLyS and these small molecule ligands could also show not only where on the BLyS surface these ligands can associate but also if they have the potential to disrupt the 60-mer complex and prevent its reformation.

Such study would offer up further points of enquiry which could allow the BLyS/TACI and BLyS/BCMA pathways to be targeted and their mechanisms could be further elucidated.

8.8 References

1. Ramanujam, M., Wang, X., Huang, W., Liu, Z., Schiffer, L., Tao, H., Frank, D., Rice, J., Diamond, B., Yu, K.O.A., Porcelli, t., and Davidson, A., *Similarities and differences between selective and nonselective BAFF blockade in murine SLE.* The Journal of Clinical Investigation, 2006. **116**(3): p. 724-734.

2. Navarra, S.V., Guzmán, R.M., Gallacher, A.E., Hall, S., Levy, R.A., Jimenez, R.E., Li, E.K.M., Thomas, M., Kim, H.-Y., León, M.G., Tanasescu, C., Nasonov, E., Lan, J.-L., Pineda, L., Zhong, Z.J., Freimuth, W., and Petri, M.A., *Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial.* The Lancet, 2011. **377**(9767): p. 721-731.

3. Mackay, F. and Mackay, C.R., *The role of BAFF in B-cell maturation, T-cell activation and autoimmunity.* Trends in Immunology, 2002. **23**(3): p. 113-115.

4. Renaudineau, Y., Garaud, S., Le Dantec, C., Alonso-Ramirez, R., Daridon, C., and Youinou, P., *Autoreactive B Cells and Epigenetics.* Clinical Reviews in Allergy and Immunology, 2010. **39**(1): p. 85-94.

5. Kim, H.M., Yu, K.S., Lee, M.E., Shin, D.R., Kim, Y.S., Paik, S.-G., Yoo, O.J., Lee, H., and Lee, J.-O., *Crystal structure of the BAFF-BAFF-R complex and its implications for receptor activation.* Nat Struct Mol Biol, 2003. **10**(5): p. 342-348.

6. Cachero, T.G., Schwartz, I.M., Qian, F., Day, E.S., Bossen, C., Ingold, K., Tardivel, A., Krushinskie, D., Eldredge, J., Silvian, L., Lugovskoy, A., Farrington, G.K., Strauch, K., Schneider, P., and Whitty, A., *Formation of Virus-like Clusters Is an Intrinsic Property of the Tumor Necrosis Factor Family Member BAFF (B Cell Activating Factor)â€* Biochemistry, 2006. **45**(7): p. 2006-2013.

7. Bodmer, J.-L., Schneider, P., and Tschopp, J.r., *The molecular architecture of the TNF superfamily.* Trends in Biochemical Sciences, 2002. **27**(1): p. 19-26.

8. Kayagaki, N., Yan, M., Seshasayee, D., Wang, H., Lee, W., French, D.M., Grewal, I.S., Cochran, A.G., Gordon, N.C., Yin, J., Starovasnik, M.A., and Dixit, V.M., *BAFF/BLyS Receptor 3 Binds the B Cell Survival Factor BAFF Ligand through a Discrete Surface Loop and Promotes Processing of NF-κB2.* Immunity, 2002. **17**(4): p. 515-524.

9. Karpusas, M., Cachero, T.G., Qian, F., Boriack-Sjodin, A., Mullen, C., Strauch, K., Hsu, Y.-M., and Kalled, S.L., *Crystal structure of extracellular*

human BAFF, a TNF family member that stimulates B lymphocytes. Journal of Molecular Biology, 2002. **315**(5): p. 1145-1154.

10. Cao, P., Mei, J.J., Diao, Z.Y., and Zhang, S.q., *Expression, refolding, and characterization of human soluble BAFF synthesized in Escherichia coli.* Protein Expression and Purification, 2005. **41**(1): p. 199-206.

11. Xue, X., Wang, Z., Yan, Z., Shi, J., Han, W., and Zhang, Y., *Production and purification of recombinant human BLyS mutant from inclusion bodies.* Protein Expression and Purification, 2005. **42**(1): p. 194-199.

12. Jungbauer, A. and Kaar, W., *Current status of technical protein refolding.* Journal of Biotechnology, 2007. **128**(3): p. 587-596.

13. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*. 3rd ed2006: Springer.

14. Liu, Y., Xu, L., Opalka, N., Kappler, J., Shu, H.-B., and Zhang, G., *Crystal Structure of sTALL-1 Reveals a Virus-like Assembly of TNF Family Ligands.* Cell, 2002. **108**(3): p. 383-394.

15. Pérot, S., Sperandio, O., Miteva, M.A., Camproux, A.-C., and Villoutreix, B.O., *Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery.* Drug Discovery Today, 2010. **15**(15-16): p. 656-667.

16. Yu, J., Zhou, Y., Tanaka, I., and Yao, M., *Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere.* Bioinformatics, 2010. **26**(1): p. 46-52.

17. Tan, K.P., Varadarajan, R., and Madhusudhan, M.S., *DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins.* Nucleic Acids Research, 2011. **39**(suppl 2): p. W242-W248.

18. Kalidas, Y. and Chandra, N., *PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins.* Journal of Structural Biology, 2008. **161**(1): p. 31-42.

19. Mehio, W., *Computational approaches for identifying inhibitors of protein interactions*, in *School of Biological Sciences*2010, University of Edinburgh: Edinburgh.

20. Hirschmann, R., Nicolaou, K.C., Pietranico, S., Leahy, E.M., Salvino, J., Arison, B., Cichy, M.A., Spoors, P.G., and Shakespeare, W.C., *De novo design and synthesis of somatostatin non-peptide peptidomimetics utilizing .beta.-D-glucose as a novel scaffolding.* Journal of the American Chemical Society, 1993. **115**(26): p. 12550-12568.

21. Thanh Le, G., Abbenante, G., Becker, B., Grathwohl, M., Halliday, J., Tometzki, G., Zuegg, J., and Meutermans, W., *Molecular diversity through sugar scaffolds.* Drug Discovery Today, 2003. **8**(15): p. 701-709.

22. Magnani, J. and Ernst, B., *Glycomimetic drugs - a new source of therapeutic opportunities.* Discov Med, 2009. **8**(43): p. 247-252.

23. Oneyama, C., Agatsuma, T., Kanda, Y., Nakano, H., Sharma, S.V., Nakano, S., Narazaki, F., and Tatsuta, K., *Synthetic Inhibitors of Proline-Rich Ligand-Mediated Protein-Protein Interaction: Potent Analogs of UCS15A.* Chemistry & Biology, 2003. **10**(5): p. 443-451.

24. Galan, M.C., Benito-Alifonso, D., and Watt, G.M., *Carbohydrate chemistry in drug discovery.* Organic & Biomolecular Chemistry. **9**(10): p. 3598-3610.

25. Walsh, J.S., Reese, M.J., and Thurmond, L.M., *The metabolic activation of abacavir by human liver cytosol and expressed human alcohol dehydrogenase isozymes.* Chemico-Biological Interactions, 2002. **142**(1-2): p. 135-154.

26. Moore, M.J., *Clinical pharmacokinetics of cyclophosphamide.* Clinical pharmacokinetics, 1991. **20**(3): p. 194-208.

27. Zalupski, M. and Baker, L.H., *Ifosamide.* JNCI J Natl Cancer Inst 1988. **80**(8): p. 556-566.

28. Varghese, A.J. and Whitmore, G.F., *Detection of a Reactive Metabolite of Misonidazole in Hypoxic Mammalian Cells.* Radiation Research, 1984. **97**(2): p. 262-271.